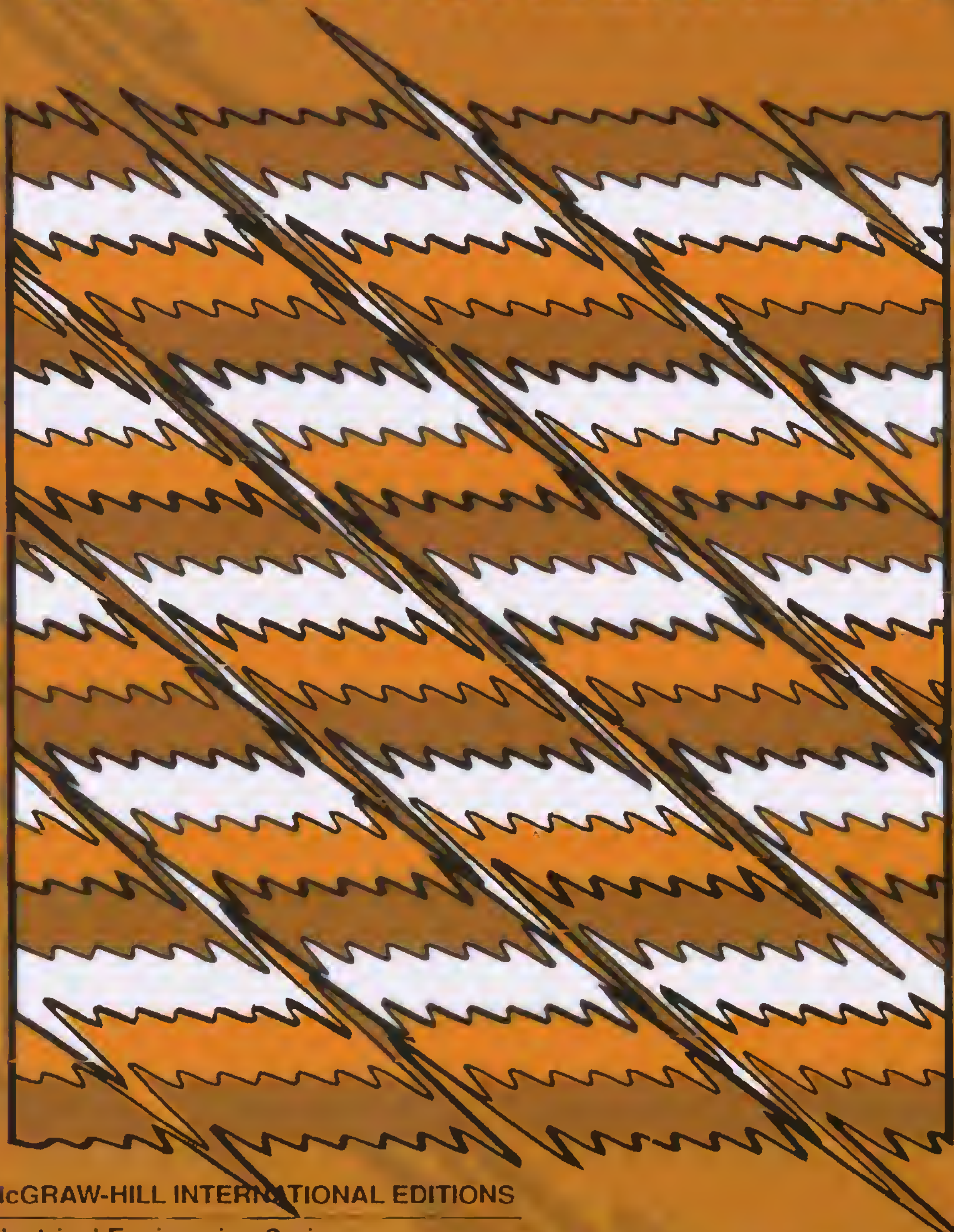
THIRD EDITION

THE FOURIER TRANSFORM AND ITS APPLICATIONS



McGRAW-HILL INTERNATIONAL EDITIONS
Electrical Engineering Series

RONALD N. BRACEWELL

Joseph Fourier, 21 March 1768–16 May 1830. (***By permission of the Bibliothèque Municipale de Grenoble.***)

The Fourier Transform and Its Applications

Third Edition

Ronald N. Bracewell
Lewis M. Terman Professor of Electrical Engineering Emeritus
Stanford University

Boston Burr Ridge, IL Dubuque, IA Madison, WI New York San Francisco St. Louis
Bangkok Bogotá Caracas Lisbon London Madrid
Mexico City Milan New Delhi Seoul Singapore Sydney Taipei Toronto

McGraw-Hill Higher Education

A Division of The McGraw-Hill Companies

THE FOURIER TRANSFORM AND ITS APPLICATIONS
International Editions 2000

2 3 4 5 6 7 8 9 0 KKP UPE 2 0

Library of Congress Cataloging-in-Publication Data

Bracewell, Ronald Newbold (date)
The Fourier transform and its applications / Roanld N. Bracewell.
– 3rd ed.
p. cm.
ISBN 0-07-303938-1
1. Fourier transformations. 2. Transformation (Mathematics).
3. Harmonic analysis. I. Title.
QA403.5.B7 2000
515/.723–dc21 99-21139

www.mhhc.com

When ordering this title, use ISBN 0-07-116043-4

Printed in Singapore

ABOUT THE AUTHOR

RONALD N. BRACEWELL was born in Australia, received his B.Sc., B.E., and M.E. degrees from the University of Sydney, and earned a Ph.D. in physics from Cambridge University. Currently L. M. Terman Professor of Electrical Engineering Emeritus at Stanford University, Dr. Bracewell has an impressive roster of professional affiliations, awards, and publications to his credit. He is a Fellow of the Royal Astronomical Society, the Astronomical Society of Australia, and past Councilor of the American Astronomical Society. He is also a life Fellow and Heinrich Hertz gold medalist of the Institute of Electrical and Electronic Engineers. At Stanford Radio Astronomy Institute he designed and built innovative radio telescopes, including the first antenna with the resolution of the human eye, less than one minute of arc, and was involved in early discoveries relating to the cosmic background radiation. Fourier analysis played a key role in his novel instrument design and data processing. Fourier's admirable ideas also contributed to Dr. Bracewell's advances in tomographic imaging which led to his election to the Institute of Medicine of the National Academy of Sciences, to receiving Sydney University's inaugural Alumni Award for Achievement, and to being appointed an Officer of the Order of Australia for service to science in the fields of radio astronomy and image reconstruction.

Fourier's theorem is not only one of the most beautiful results of modern analysis, but it may be said to furnish an indispensable instrument in the treatment of nearly every recondite question in modern physics.
Lord Kelvin

To my wife Helen,
whose support made
this edition possible.
Ronald Bracewell

CONTENTS

PREFACE

Transform methods provide a unifying mathematical approach to the study of electrical networks, devices for energy conversion and control, antennas, and other components of electrical systems, as well as to complete linear systems and to many other physical systems and devices, whether electrical or not. These same methods apply equally to the subjects of electrical communication by wire or optical fiber, to wireless radio propagation, and to ionized media—which are all concerned with the interconnection of electrical systems—and to information theory which, among other things, relates to the acquisition, processing, and presentation of data. Other theoretical techniques are used in handling these basic fields of electrical engineering, but transform methods are virtually indispensable in all of them. Fourier analysis as applied to electrical engineering is sufficiently important to have earned a permanent place in the curriculum—indeed much of the mathematical development took place in connection with alternating current theory, signal analysis, and information theory as formulated in connection with electrical communication.

This is why much of the literature dealing with technical applications has appeared in electrical and electronic journals. Despite the strong bonds with electrical engineering, Fourier analysis nevertheless has become indispensable in biomedicine and remote sensing (geophysics, oceanography, planetary surfaces, civil engineering), where practitioners now outnumber those electrical engineers who regularly use Fourier analysis. But the teaching of Fourier analysis and its applications still finds its home in electrical engineering.

A course on transforms and their applications has formed part of the electrical engineering curriculum at Stanford University for many years and has been given with no prerequisites beyond those that the holder of a bachelor's degree normally possesses. One objective has been to develop a pivotal course to be taken at an early stage by all graduates, so that in later, more specialized courses, the student would be spared encountering the same material over and over again; later instructors can then proceed more directly to their special subject matter.

It is clearly not feasible to give the whole of linear mathematics in a single course; the choice of core material must necessarily remain a matter of local judgment. The choice will, however, be of most help to later instructors if sharply defined.

An early-level course should be simple, but not trivial; the objective of this book is to simplify the presentation of many key topics that are ordinarily dealt

with in advanced contexts, by making use of suitable notation and an approach through convolution.

One way of working from the book is to begin by taking the chapters in numerical order. This sequence is feasible for students who could read the first half unassisted or who could be taken through it rapidly in a few lectures; but if the material is approached at a more normal pace, then, as a practical matter, it is a good idea to interpret each theorem and concept in terms of a physical example. Waveforms and their spectra and elementary diffraction theory are suitable. After that, the chapters on applications can then be selected in whatever sequence is desired. The organization of chapters is as follows:

1. Introduction
↓
2. Groundwork
↓
3. Convolution
↓
4. Notation . . .
↓
5. The Impulse Symbol
↓
6. The Basic Theorems
↓
7. Obtaining Transforms
↓
8. The Two Domains →

Choices →
- 9. Waveforms, Spectra, . . .
- 10. Sampling and Series
- 11. Discrete FT and FFT
- 12. Hartley Transform
- 13. Relatives → 14. Laplace

Applications →
- 15. Antennas and Optics
- 19. Dynamic Spectra
- 16. Statistics → 17. Noise
- 18. Heat and Diffusion

Reference →
- 20. Tables of sinc, . . .
- 21. Solutions to Problems
- 22. Pictorial Dictionary
- 23. Biography of Fourier

The amount of material is suitable for one semester, or for one quarter, according to how many of the later chapters on applications are included. A practical plan is to leave the choice of chapters on applications to the current instructor.

Many fine mathematical texts on the Fourier transform have been published. This book differs in that it is intended for those who are concerned with applying Fourier transforms to physical situations rather than with pursuing the mathematical subject as such. The connections of the Fourier transform with other transforms are also explored, and the text has been purposely enriched with condensed information that will suit it for use as a reference source for transform pairs and theorems pertaining to transforms.

My interest in the subject was fired while studying analysis from H. S. Carslaw's "Fourier Series and Integrals" at the University of Sydney in 1939. I learned about physical applications as a colleague of J. C. Jaeger at C.S.I.R Radiophysics Laboratory and inherited the physical wisdom of the crystallographers of the Cavendish Laboratory, Cambridge, as transmitted by J. A. Ratcliffe. Transform methods are at the heart of the electrical engineering curriculum. Digital

computing and data processing, which have emerged as large curricular segments, though rather different in content from the rigorous study of circuits, electronics, and waves, nevertheless do share a common bond through the Fourier transform. The diffusion equation, which long ago had a connection with submarine cable telegraphy, has reemerged as an essential consideration in solid state physics and devices, both through the practice of doping, by which semiconductor devices are fabricated, and as a controlling influence in electrical conduction by holes and electrons. Needless to say, a grasp of Fourier fundamentals is an asset in the solid-state laboratory.

The explosion of image engineering, much of which can be interpreted via two-dimensional generalization, has reinforced the value of a core course. Consequently, the subject matter of this book has easily moved into the pivotal role foreseen for it, and faculty members from various specialties have found themselves comfortable teaching it. The course is taken by first-year graduate students, especially students arriving from other universities, and by students from other departments, notably applied physics and earth sciences. The course is accessible to students in the last year of their bachelor's degree.

Introduction of the fast Fourier transform (FFT) algorithm has greatly broadened the scope of application of the Fourier transform to data handling and to digital formulation in general and has brought prominence to the discrete Fourier transform (DFT). The technological revolution associated with discrete mathematics as treated in Chapter 11 has made an understanding of Fourier notions (such as aliasing, which only aficionados used to guard against) indispensable to any professional who handles masses of data, not only engineers but experts in many subfields of medicine, biology, and remote sensing. Developments based on Ralph V. L. Hartley's equations (Chapter 12) have made it possible to dispense with imaginaries in computed Fourier analysis and to proceed elegantly and simply using the real Hartley formalism.

Hartley's equations, which quietly received honorable mention in the first edition of this book, gained major relevance to signal processing as computers flourished. In 1983 I gave them new life in a time-series context with modern notation under the title of discrete Hartley transform, a name that is now universally recognized, while Z. Wang (*Appl. Math. and Comput.*, vol. 9, pp. 53–73, 153–163, 245–255, 1983) independently stimulated mathematicians. Hartley's cas (*co*sine *a*nd *s*ine) function is now widely recognized.

For those who like to do their own computer programming some segments of pseudocode have been supplied. Translating into your language of choice may give some insights that complement the algebraic and graphical viewpoints. Furthermore, executing numerical examples develops a useful sort of intuition which, while not as powerful as physical intuition, adds a further dimension to one's experience. Pseudocode is suited to readers who cannot be expected to know several popular languages. The aim is to provide the simplest intelligible instructions that are susceptible to simultaneous transcription into the language of fluency of the reader, who provides the necessary protocol, array declarations, and other distinctive features.

The code segments in this book are presented to supplement verbal explanation, not to be a substitute for a computational toolbox.

However, it is often more important to be able to use a computer algorithm than to understand in detail how it was constructed, just as when using a table of integrals or an engineering design handbook. To meet this need and to bring the power of Fourier transformation into the hands of a much wider constituency, packages of software tools have been created commercially and have become indispensable. A popular example is MATLAB®, a user-friendly, higher-level, special-purpose application whose use is illustrated in Chapters 7 and 11.

Caution is needed in circumstances where the user is shielded from the algorithmic details; it is handy to know what to expect before being presented with computer output. For this and other reasons, transforms presented graphically in the Pictorial Dictionary have proved to be a useful reference feature. Graphical presentation is a useful adjunct to the published compilations of integral transforms, where it is sometimes frustrating to seek commonly needed entries among the profusion of rare cases and where, in addition, simple functions that are impulsive, discontinuous, or defined piecewise may be hard to recognize or may not be included at all.

A good problem assigned at the right stage can be extremely valuable for the student, but a good problem is hard to compose. Many of the problems here go beyond mathematical exercises by inclusion of technical background or by asking for opinions. Those wishing to mine the good material in the problems will appreciate that many of them are now titled, a practice that should be more widely adopted. Many of the problems are discussed in Chapter 21, but occasionally it is nice to have a new topic followed by an exercise that is in close proximity rather than at the end of the chapter; a sprinkling of these is provided.

Notation is a vital adjunct to thinking, and I am happy to report that the *sinc function*, which we learned from P. M. Woodward, is alive and well, and surviving erosion by occasional authors who do not know that "sine x over x" is not the sinc function. The *unit rectangle function* (unit height and width) $\Pi(x)$, the transform of the sinc function, has also proved extremely useful, especially for blackboard work. In typescript or other media where the Greek letter is less desirable, $\Pi(x)$ may be written "rect x," and it is convenient in any case to pronounce it *rect*. The *jinc function*, the circular analogue of the sinc function, has the corresponding virtues of normalization and the distinction of describing the diffraction field of a telescope or camera. The *shah function* $\text{III}(x)$ has caught on. It is easy to print and is twice as useful as you might think because it is its own transform. The asterisk for convolution, which was in use a long time ago by Volterra and perhaps earlier, is now in wide use and I recommend ** to denote two-dimensional convolution, which has become common as a result of the explosive growth of image processing.

Early emphasis on digital convolution in a text on the Fourier transform turned out to be exactly the way to start. Convolution has changed in a few years from being presented as a rather advanced concept to one that can be easily explained at an early stage, as is fitting for an operation that applies to all those systems that respond sinusoidally when you shake them sinusoidally.

1

Introduction

Linear transforms, especially those named for Fourier and Laplace, are well known as providing techniques for solving problems in linear systems. Characteristically one uses the transformation as a mathematical or physical tool to alter the problem into one that can be solved. This book is intended as a guide to the understanding and use of transform methods in dealing with linear systems.

The subject is approached through the Fourier transform. Hence, when the more general Laplace transform is discussed later, many of its properties will already be familiar and will not distract from the new and essential question of the strip of convergence on the complex plane. In fact, all the other transforms discussed here are greatly illuminated by an approach through the Fourier transform.

Fourier transforms play an important part in the theory of many branches of science. While they may be regarded as purely mathematical functionals, as is customary in the treatment of other transforms, they also assume in many fields just as definite a physical meaning as the functions from which they stem. A waveform—optical, electrical, or acoustical—and its spectrum are appreciated equally as physically picturable and measurable entities: an oscilloscope enables us to see an electrical waveform, and a spectroscope or spectrum analyzer enables us to see optical or electrical spectra. Our acoustical appreciation is even more direct, since the ear hears spectra. Waveforms and spectra are Fourier transforms of each other; the Fourier transformation is thus an eminently physical relationship.

The number of fields in which Fourier transforms appear is surprising. It is a common experience to encounter a concept familiar from one branch of study in a slightly different guise in another. For example, the principle of the phase-contrast microscope is reminiscent of the circuit for detecting frequency modulation, and the explanation of both is conveniently given in terms of transforms along the same lines. Or a problem in statistics may yield to an approach which is familiar from studies of cascaded amplifiers. This is simply a case of the one underlying theorem from Fourier theory assuming different physical embodiments.

It is a great advantage to be able to move from one physical field to another and to carry over the experience already gained, but it is necessary to have the key which interprets the terminology of the new field. It will be evident, from the rich variety of topics coming within its scope, what a pervasive and versatile tool Fourier theory is.

Many scientists know Fourier theory not in terms of mathematics, but as a set of propositions about physical phenomena. Often the physical counterpart of a theorem is a physically obvious fact, and this allows the scientist to be abreast of matters which in the mathematical theory may be quite abstruse. Emphasis on the physical interpretation enables us to deal in an elementary manner with topics which in the normal course of events would be considered advanced.

Although the Fourier transform is vital in so many fields, it is often encountered in formal mathematics courses in the last lecture of a formidable course on Fourier series. As need arises it is introduced ad hoc in later graduate courses but may never develop into a usable tool. If this traditional order of presentation is reversed, the Fourier series then falls into place as an extreme case within the framework of Fourier transform theory, and the special mathematical difficulties with the series are seen to be associated with their extreme nature—which is nonphysical; the handicap imposed on the study of Fourier transforms by the customary approach is thus relieved.

The great generality of Fourier transform methods strongly qualifies the subject for introduction at an early stage, and experience shows that it is quite possible to teach, at this stage, the distilled theorems, which in their diverse applications offer such powerful tools for thinking out physical problems.

The present work began as a pictorial guide to Fourier transforms to complement the standard lists of pairs of transforms expressed mathematically. It quickly became apparent that the commentary would far outweigh the pictorial list in value, but the pictorial dictionary of transforms is nevertheless important, for a study of the entries reinforces the intuition, and many valuable and common types of function are included which, because of their awkwardness when expressed algebraically, do not occur in other lists.

A contribution has been made to the handling of simple but awkward functions by the introduction of compact notation for a few basic functions which are usually defined piecewise. For example, the rectangular pulse, which is at least as simple as a Gaussian pulse, is given the name $\Pi(x)$, which means that it can be handled as a simple function. The picturesque term "gate function," which is in use in electronics, suggests how a gating waveform $\Pi(t)$ opens a valve to let through a segment of a waveform. This is the way we think of the rectangle function mathematically when we use it as a multiplying factor.

Among the special symbols introduced or borrowed are $\text{II}(x)$, the even impulse pair, and $\text{III}(x)$ (pronounced *shah*), the infinite impulse train defined by $\text{III}(x) = \Sigma\, \delta(x - n)$. The first of these two gains importance from its status as the Fourier transform of the cosine function; the second proves indispensable in discussing both regular sampling or tabulation (operations which are equivalent to multiplication by *shah*) and periodic functions (which are expressible as convolu-

tions with *shah*). Since *shah* proves to be its own Fourier transform, it turns out to be twice as useful an entity as might have been expected. Much freedom of expression is gained by the use of these conventions of notation, especially in conjunction with the asterisk notation for convolution. Only a small step is involved in writing $\Pi(x) * f(x)$, or simply $\Pi * f$, instead of

$$\int_{x-\frac{1}{2}}^{x+\frac{1}{2}} f(u)\, du$$

(for example, for the response to a photographic density distribution $f(x)$ on a sound track scanned with a slit). But the disappearance of the dummy variable and the integral sign with limits, and the emergence of the character of the response as a convolution between two profiles Π and f, lead to worthwhile convenience in both algebraic and mental manipulation.

Convolution is used a lot here. Experience shows that it is a fairly tricky concept when it is presented bluntly under its integral definition, but it becomes easy if the concept of a functional is first understood. Numerical practice on serial products confirms the feeling for convolution and incidentally draws attention to the practical character of numerical evaluation: for numerical purposes one normally prefers to have the answer to a problem come out as the convolution of two functions rather than as a Fourier transform.

This is a good place to mention that transform methods do not necessarily involve taking transforms numerically. On the contrary, some of the best methods for handling linear problems do not involve application of the Fourier or Laplace transform to the data at all; but the basis for such methods is often clarified by appeal to the transform domain. Thinking in terms of transforms, we may show how to avoid numerical harmonic analysis or the handling of data on the complex plane.

It is well known that the response of a system to harmonic input is itself harmonic, at the same frequency, under two conditions: linearity and time invariance of the system properties. These conditions are, of course, often met. This is why Fourier analysis is important, why one specifies an amplifier by its *frequency* response, why harmonic variation is ubiquitous. When the conditions for harmonic response to harmonic stimulus break down, as they do in a nonlinear servomechanism, analysis of a stimulus into harmonic components must be reconsidered. *Time* invariance can often be counted on even when linearity fails, but *space* invariance is by no means as common. Failure of this condition is the reason that bridge deflections are not studied by analyzing the load distribution into sinusoidal components (space harmonics).

The two conditions for harmonic response to harmonic stimulus can be restated as *one* condition: that the response shall be relatable to the stimulus by *convolution*. For work in Fourier analysis, convolution is consequently profoundly important, and such a pervasive phenomenon as convolution does not lack for familiar examples. Good ones are the relation between the distribution on a film sound track or recorder tape and the electrical signal read out through the scanning slit or magnetic head.

Many topics normally considered abstruse or advanced are presented here, and simplification of their presentation is accomplished by minor conveniences of notation and by the use of graphs.

Special care has been given to the presentation and use of the impulse symbol $\delta(x)$, on which, for example, both $\mathrm{II}(x)$ and $\mathrm{III}(x)$ depend. The term "impulse symbol" focuses attention on the status of $\delta(x)$ as something which is not a function; equations or expressions containing it then have to have an interpretation, and this is given in an elementary fashion by recourse to a sequence of pulses (not impulses). The expression containing the impulse symbol thus acquires meaning as a limit which, in many instances, "exists." This commonplace mathematical status of the *complete expression*, in contrast with that of $\delta(x)$ itself, directly reflects the physical situation, where Green's functions, impulse responses, and the like are often accurately producible or observable, within the limits permitted by the resolving power of the measuring equipment, while impulses themselves are fictitious. By deeming all expressions containing $\delta(x)$ to be subject to special rules, we can retain both rigor and the direct procedures of manipulating $\delta(x)$ which have been so successful.

Familiar examples of this physical situation are the moment produced by a point mass resting on a beam and the electric field of a point charge. In the physical approach to such matters, which have long been imbedded in physics, one thinks about the effect produced by smaller and smaller but denser and denser massive objects or charged volumes, and notes whether the effect produced approaches a definite limit. Ways of representing this mathematically have been tidied up to the satisfaction of mathematicians only in recent years, and use of $\delta(x)$ is now licensed if accompanied by an appropriate footnote reference to this recent literature, although in some conservative fields such as statistics $\delta(x)$ is still occasionally avoided in favor of distinctly more awkward Stieltjes integral notation. These developments in mathematical ideas are mentioned in Chapters 5 and 6, the presentation in terms of "generalized functions" following Temple and Lighthill being preferred. The validity of the original physical ideas remains unaffected, and one ought to be able, for example, to discuss the moment of a point mass on a beam by considering those rectangular distributions of pressure that result from restacking the load uniformly on an ever-narrower base to an ever-increasing height; it should not be necessary to limit attention to pressure distributions possessing an infinite number of continuous derivatives merely because in some other problem a derivative of high order is involved. Therefore the subject of impulses is introduced with the aid of the rather simple mathematics of rectangle functions, and in the relatively few cases where the first derivative is wanted, triangle functions are used instead.

2

Groundwork

Most of the material in this chapter is stated without proof. This is done because the proofs entail discussions that are lengthy (in fact, they form the bulk of conventional studies in Fourier theory) and remote from the subject matter of the present work.

Omitting the proofs enables us to take the transform formulas and their known conditions as our point of departure. Since suitable notation is an important part of the work, it too is set out in this chapter.

THE FOURIER TRANSFORM AND FOURIER'S INTEGRAL THEOREM

The Fourier transform of $f(x)$ is defined as

$$\int_{-\infty}^{\infty} f(x)e^{-i2\pi xs}\,dx.$$

This integral, which is a function of s, may be written $F(s)$. Transforming $F(s)$ by the same formula, we have

$$\int_{-\infty}^{\infty} f(s)e^{-i2\pi ws}\,ds.$$

When $F(x)$ is an even function of x, that is, when $f(x) = f(-x)$, the repeated transformation yields $f(w)$, the same function we began with. This is the cyclical property of the Fourier transformation, and since the cycle is of two steps, the reciprocal property is implied: if $F(s)$ is the Fourier transform of $f(x)$, then $f(x)$ is the Fourier transform of $F(s)$.

The cyclical and reciprocal properties are imperfect, however, because when $f(x)$ is odd—that is, when $f(x) = -f(-x)$—the repeated transformation yields

$f(-w)$. In general, whether $f(x)$ is even or odd or neither, repeated transformation yields $f(-w)$.

The customary formulas exhibiting the reversibility of the Fourier transformation are

$$F(s) = \int_{-\infty}^{\infty} f(x)e^{-i2\pi xs}\,dx$$

$$f(x) = \int_{-\infty}^{\infty} F(s)e^{i2\pi xs}\,ds.$$

In this form, two successive transformations are made to yield the original function. The second transformation, however, is not exactly the same as the first, and where it is necessary to distinguish between these two sorts of Fourier transform, we shall say that $F(s)$ is the minus-i transform of $f(x)$ and that $f(x)$ is the plus-i transform of $F(s)$.

Writing the two successive transformations as a repeated integral, we obtain the usual statement of Fourier's integral theorem:

$$f(x) = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} f(x)e^{-i2\pi xs}\,dx\right] e^{i2\pi xs}\,ds.$$

The conditions under which this is true are given in the next section, but it must be stated at once that where $f(x)$ is discontinuous the left-hand side should be replaced by $\frac{1}{2}[f(x+) + f(x-)]$, that is, by the mean of the unequal limits of $f(x)$ as x is approached from above and below.

The factor 2π appearing in the transform formulas may be lumped with s to yield the following version (system 2):

$$F(s) = \int_{-\infty}^{\infty} f(x)e^{-ixs}\,dx$$

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(s)e^{ixs}\,ds.$$

And for the sake of symmetry, authors occasionally write (system 3):

$$F(s) = \frac{1}{(2\pi)^{\frac{1}{2}}}\int_{-\infty}^{\infty} f(x)e^{-ixs}\,dx$$

$$f(x) = \frac{1}{(2\pi)^{\frac{1}{2}}}\int_{-\infty}^{\infty} F(s)e^{ixs}\,ds.$$

All three versions are in common use, but here we shall keep the 2π in the exponent (system 1). If $f(x)$ and $F(s)$ are a transform pair in system 1, then $f(x)$ and $F(s/2\pi)$ are a transform pair in system 2, and $[x/2\pi)^{\frac{1}{2}}]$ and $F(s/2\pi)^{\frac{1}{2}}]$ are a transform pair in system 3. An example of a transform pair in each of the three systems follows.

System 1		System 2		System 3	
$f(x)$	$F(s)$	$f(x)$	$F(s)$	$f(x)$	$F(s)$
$e^{-\pi x^2}$	$e^{-\pi s^2}$	$e^{-\pi x^2}$	$e^{-s^2/4\pi}$	$e^{-\frac{1}{2}x^2}$	$e^{-\frac{1}{2}s^2}$

An excellent notation which may be used as an alternative to $F(s)$ is $\overline{f}(s)$. Various advantages and disadvantages are found in both notations. The bar notation leads to compact expression, including some convenient manuscript or blackboard forms which are not very suitable for typesetting. Consider, for example, these versions of the convolution theorem,

$$\overline{FG} = \overline{F} * \overline{G}$$
$$\overline{F * G} = \overline{F}\,\overline{G}$$
$$\overline{\overline{F}\,\overline{G}} = F * G$$
$$\overline{\overline{F} * \overline{G}} = FG,$$

which display shades of distinction not expressible at all in the capital notation. See Chapter 6 for illustrations of the freedom of expression permitted by the bar notation. A certain awkwardness sets in, however, when complex conjugates or primes representing derivatives have to be handled; this awkwardness does not afflict the capital notation. Therefore we have departed from the usual custom of adopting a single notation. In the early mathematical sections, where f and g are nearly the only symbols for functions, the capital notation is mainly used. In the physical sections, preemption of capitals such as E and H for the representation of physical quantities leads more naturally to bars.

Neither of the above two notations lends itself to symbolic statements equivalent to "the Fourier transform of exp $(-\pi x^2)$ is exp $(-\pi s^2)$"; however, we can write

$$\mathcal{F}e^{-\pi x^2} = e^{-\pi s^2}$$

or

$$e^{-\pi x^2} \supset e^{-\pi s^2}.$$

In the first of these, $\mathcal{F}$ can be regarded as a functional operator which converts a function into its transform. It may be applied wherever the bar and capital notations are employed, but will be found most appropriate in connection with specific functions such as those above. It lends itself to the use of affixes, (for example, $\mathcal{F}_s$ and $\mathcal{F}_c$ and ${}^2\mathcal{F}$) and mixes with symbols for other transforms. It can also distinguish between the minus-i and plus-i transforms through the use of $\mathcal{F}^{-1}$ for the inverse of $\mathcal{F}$; or like the bar notation, but not the capital notation, it can remain discreetly silent. The properties of this notation make it indispensable, and it is adopted in suitable places in the sequel.

The sign $\supset$ is not as versatile as $\mathcal{F}$ but is simple, and useful for algebraic work; it is also used to denote the Laplace transform and, for occasional use, $\subset$

is available to indicate the inverse transform. Nonreversibility is also conveyed by $\rightleftharpoons$ and its reverse, but it is not obvious which means direct and which means inverse. The signs $\rightarrow$ and $\Rightarrow$ are not recommended because they have established meanings as "approaches" and "implies" respectively. For reversible transforms such as the cosine, Hilbert and Hartley transforms the signs $\leftrightarrow$ and $\Leftrightarrow$ do have the appropriate symmetry.

CONDITIONS FOR THE EXISTENCE OF FOURIER TRANSFORMS

A circuit expert finds it obvious that every waveform has a spectrum, and the antenna designer is confident that every antenna has a radiation pattern. It sometimes comes as a surprise to those whose acquaintance with Fourier transforms is through physical experience rather than mathematics that there are some functions without Fourier transforms. Nevertheless, we may be confident that no one can generate a waveform without a spectrum or construct an antenna without a radiation pattern.

The question of the existence of transforms may safely be ignored when the function to be transformed is an accurately specified description of a physical quantity. Physical possibility is a valid sufficient condition for the existence of a transform. Sometimes, however, it is convenient to substitute a simple mathematical expression for a physical quantity. It is very common, for example, to consider the waveforms

$\sin t$	(harmonic wave, pure alternating current)
$H(t)$	(step)
$\delta(t)$	(impulse).

It turns out that none of these three has, strictly speaking, a Fourier transform. Of course, none of them is physically possible, for a waveform $\sin t$ would have to have been switched on an infinite time ago, a step $H(t)$ would have to be maintained steady for an infinite time, and an impulse $\delta(t)$ would have to be infinitely large for an infinitely short time. However, in a given situation we can often achieve an approximation so close that any further improvement would be immaterial, and we use the simple mathematical expressions because they are less cumbersome than various slightly different but realizable functions. Nevertheless, the above functions do not have Fourier transforms; that is, the Fourier integral does not converge for all s. It is therefore of practical importance to consider the conditions for the existence of transforms.

Transforming and retransforming a single-valued function $f(x)$, we have the repeated integral

$$\int_{-\infty}^{\infty} e^{i2\pi sx}\left[\int_{-\infty}^{\infty} f(x)e^{-i2\pi sx}\,dx\right] ds.$$

This expression is equal to $f(x)$ (or to $\frac{1}{2}[f(x+) + f(x-)]$ where $f(x)$ is discontinuous), provided that

1. The integral of $|f(x)|$ from $-\infty$ to ∞ exists
2. Any discontinuities in $f(x)$ are finite

A further but less important condition is mentioned below. In physical circumstances these conditions are violated[1] when there is infinite energy, and a kind of duality between the two conditions is often noted. For instance, absolutely steady direct current which has always been flowing and always will flow represents infinite energy and violates the first condition. The distribution of energy with frequency would have to show infinite energy concentrated entirely at zero frequency and would violate the second condition. The same applies to harmonic waves.

It is sometimes stated that an infinite number of maxima and minima in a finite interval disqualifies a function from possessing a Fourier transform, the stock example being $\sin x^{-1}$ (see Fig. 2.1), which oscillates with ever-increasing frequency as x approaches zero. This kind of behavior is not important in real life, even as an approximation. We therefore record for general interest that some functions with infinite numbers of maxima and minima in a finite interval do have transforms. This is allowed for when the further condition is given as *bounded variation*.[2] Again, however, there are transformable functions with an infinite number of maxima and with unbounded variation in a finite interval, a circumstance which may be covered by requiring $f(x)$ to satisfy a Lipschitz condition.[3] Then there is a more relaxed condition used by Dini. This is a fascinating topic in Fourier theory, but it is not immediately relevant to our branch of it, which is physical applications. Furthermore, we by no means propose to abandon useful functions which do not possess Fourier transforms in the ordinary sense. On the contrary, we include them equally, by means of a generalization to Fourier transforms in the limit. Conditions for the existence of Fourier transforms now merely distinguish, where distinction is desired, between those transforms which are ordinary and those which are transforms in the limit.

---

[1]Exceptions are provided by finite-energy waveforms such as $(1 + |x|)^{-3/4}$ and $x^{-1} \sin x$, which nevertheless do not have absolutely convergent infinite integrals.

[2]A function $f(x)$ has bounded variation over the interval $x = a$ to $x = b$ if there is a number M such that

$$|f(x_1) - f(a)| + |f(x_2) - f(x_1)| + \cdots + |f(b) - f(x_{n-1})| \leq M$$

for every method of subdivision $a < x_1 < x_2 < \ldots < x_{n-1} < b$. Any function having an absolutely integrable derivative will have bounded variation.

[3]A function $f(x)$ satisfies a Lipschitz condition of order α at $x = 0$ if

$$|f(h) - f(0)| \leq B|h|^{\beta}$$

for all $|h| < \epsilon$, where B and β are independent of h, β is positive, and α is the upper bound of all β for which finite B exists.

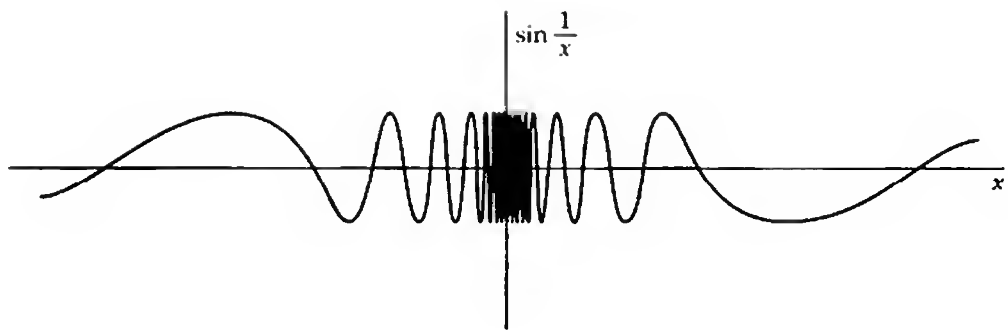


Fig. 2.1 **A function with an infinite number of maxima.**

TRANSFORMS IN THE LIMIT

Although a periodic function does not have a Fourier transform, as may be verified by reference to the conditions for existence, it is nevertheless considered in physics to have a spectrum, a "line spectrum." The line spectrum may be identified with the coefficients of the Fourier series for the periodic function, or we may broaden the mathematical concept of the Fourier transform to bring it into harmony with the physical standpoint. This is what we shall do here, taking the periodic function as one example among others which we would like Fourier transform theory to embrace.

Let $P(x)$ be a periodic function of x. Then

$$\int_{-\infty}^{\infty} |P(x)|\, dx$$

does not exist, but if we modify $P(x)$ slightly by multiplication with a factor such as $\exp(-\alpha x^2)$, where α is a small positive number, then the modified version may have a transform, for

$$\int_{-\infty}^{\infty} |e^{-\alpha x^2} P(x)|\, dx$$

may exist. Of course, any infinite discontinuities in $P(x)$ will still disqualify it, but let us select $P(x)$ so that $\exp(-\alpha x^2)P(x)$ possesses a Fourier transform. Then as α approaches zero, the modifying factor for each value of x approaches unity, and the modified functions of the sequence generated as α approaches zero thus approach $P(x)$ in the limit. Since each modified function possesses a transform, a corresponding sequence of transforms is generated; now, as α approaches zero, does this sequence of transforms also approach a limit? We already know that it does not, at least not for all s; we content ourselves with saying that the sequence of regular transforms defines or constitutes an entity that shall be called a gener-

alized function. The periodic function and the generalized function form a Fourier transform pair in the limit.

The idea of dealing with things that are not functions but are describable in terms of sequences of functions is well established in physics in connection with the impulse symbol $\delta(x)$. In this case a progression of ever-stronger and ever-narrower unit-area pulses is an appropriate sequence, and a little later in the chapter we go into this idea more fully. We use the term "generalized function" to cover impulses and their like.

Periodic functions fail to have Fourier transforms because their infinite integral is not absolutely convergent; failure may also be due to the infinite discontinuities associated with impulses. In this case we replace any impulse by a sequence of functions that do have transforms; then the sequence of corresponding transforms may approach a limit, and again we have a Fourier transform pair in the limit. As before, only one member of the pair is a generalized function involving impulses.

It may also happen that the sequence of transforms does not approach a limit. This would be so if we began with something that was both impulsive and periodic; then the members of the transform pair in the limit would both be generalized functions involving impulses.

At this point we might proceed to lay the groundwork leading to the definition of a generalized function. Instead we defer the rather severe general discussion to a much later stage, since it can be read with more profit after facility has been acquired in handling the impulse symbol $\delta(x)$.

ODDNESS AND EVENNESS

Symmetry properties play an important role in Fourier theory. Arguments from symmetry to show directly that certain integrals vanish, without the need of evaluating them, are familiar and perhaps often seem trivial in print. More alertness is needed, however, to ensure full exploitation in one's own reasoning of symmetry restrictions and the corresponding restrictive properties generated under Fourier transformation. Some simple terminology is recalled here.

A function $E(x)$ such that $E(-x) = E(x)$ is a symmetrical, or even, function. A function $O(x)$ such that $O(-x) = -O(x)$ is an antisymmetrical, or odd, function (see Fig. 2.2). The sum of even and odd functions is in general neither even nor odd, as illustrated in Fig. 2.3, which shows the sum of the previously chosen examples.

Any function $f(x)$ can be split unambiguously into odd and even parts. For if $f(x) = E_1(x) + O_1(x) = E_2(x) + O_2(x)$, then $E_1 - E_2 = O_2 - O_1$; but $E_1 - E_2$ is even and $O_2 - O_1$ odd, hence $E_1 - E_2$ must be zero.

The even part of a given function is the mean of the function and its reflection in the vertical axis, and the odd part is the mean of the function and its neg-

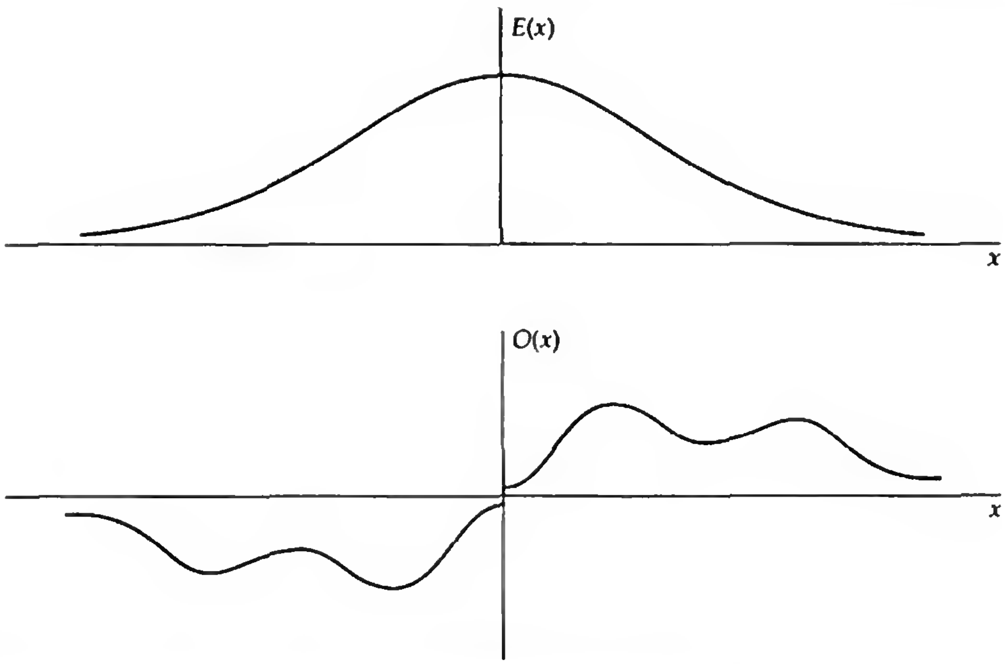


Fig. 2.2 An even function $E(x)$ and an odd function $O(x)$.

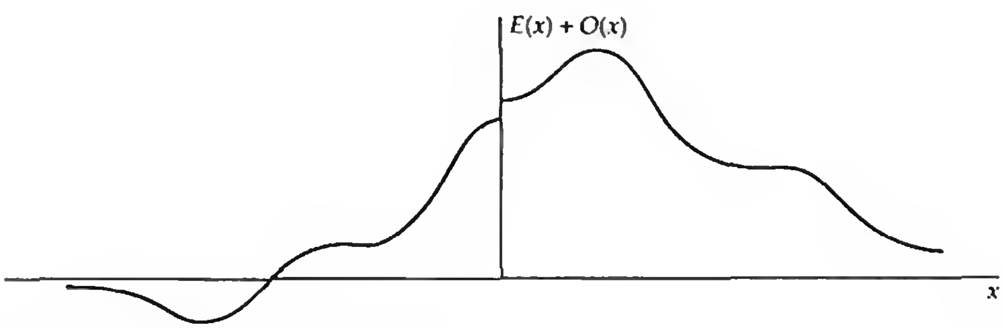


Fig. 2.3 The sum of $E(x)$ and $O(x)$.

ative reflection (see Fig. 2.4). Thus

$$E(x) = \tfrac{1}{2}[f(x) + f(-x)]$$

and

$$O(x) = \tfrac{1}{2}[f(x) - f(-x)].$$

The dissociation into odd and even parts changes with changing origin of x, some functions such as $\cos x$ being convertible from fully even to fully odd by a shift of origin.

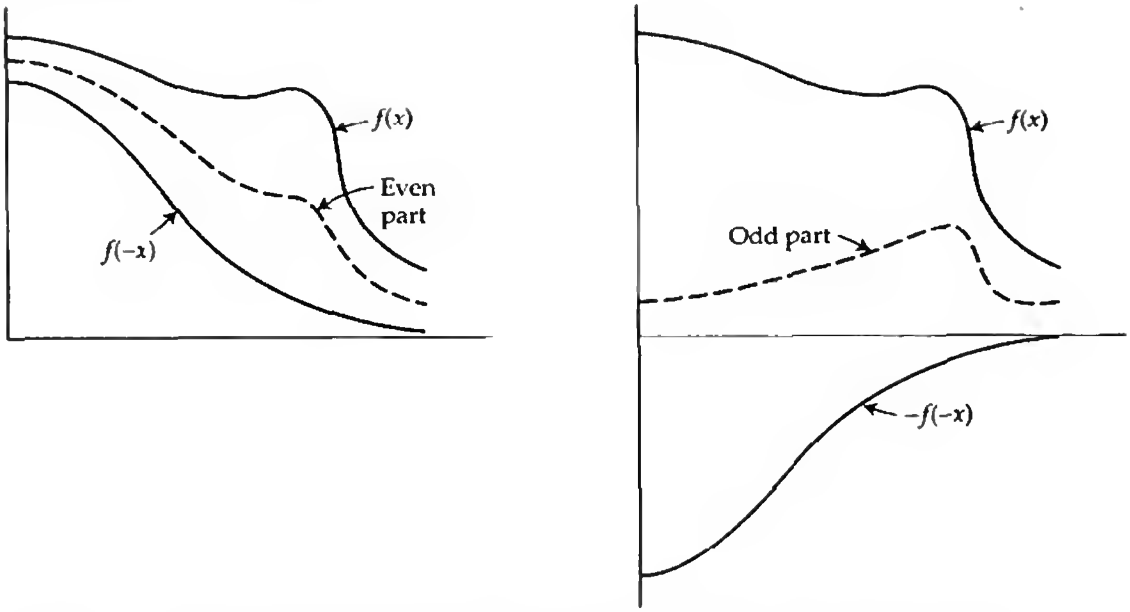


Fig. 2.4 **Constructions for the even and odd parts of a given function $f(x)$.**

SIGNIFICANCE OF ODDNESS AND EVENNESS

Let

$$f(x) = E(x) + O(x),$$

where E and O are in general complex. Then the Fourier transform of $f(x)$ reduces to

$$2\int_0^\infty E(x)\cos(2\pi xs)\,dx - 2i\int_0^\infty O(x)\sin(2\pi xs)\,dx.$$

It follows that if a function is even, its transform is even, and if it is odd, its transform is odd. Full results are

Real and even	Real and even
Real and odd	Imaginary and odd
Imaginary and even	Imaginary and even
Complex and even	Complex and even
Complex and odd	Complex and odd
Real and asymmetrical	Complex and hermitian
Imaginary and asymmetrical	Complex and antihermitian
Real even plus imaginary odd	Real
Real odd plus imaginary even	Imaginary
Even	Even
Odd	Odd

These properties are summarized in the following diagram:

$$
\begin{tikzcd}[column sep=tiny]
f(x) & = & o(x) \arrow[d] & + & e(x) \arrow[d] & = & \mathrm{Re}\, o(x) \arrow[drr] & + & i\, \mathrm{Im}\, o(x) \arrow[dll] & + & \mathrm{Re}\, e(x) \arrow[d] & + & i\, \mathrm{Im}\, e(x) \arrow[d] \\
F(s) & = & O(s) & + & E(s) & = & \mathrm{Re}\, O(s) & + & i\, \mathrm{Im}\, O(s) & + & \mathrm{Re}\, E(s) & + & i\, \mathrm{Im}\, E(s).
\end{tikzcd}
$$

Figure 2.5, which records the phenomena in another way, is also valuable for revealing at a glance the "relative sense of oddness": when $f(x)$ is real and odd with a *positive* moment, the odd part of $F(s)$ has i times a *negative* moment; and when $f(x)$ is real but not necessarily odd, we also find opposite senses of oddness. However, inverting the procedure—that is, going from $F(s)$ to $f(x)$, or taking $f(x)$ to be imaginary—produces the same sense of oddness.

Real even functions play a special part in this work because both they and their transforms may easily be graphed. Imaginary odd, real odd, and imaginary even functions are also important in this respect.

Another special kind of symmetry is possessed by a function $f(x)$ whose real part is even and imaginary part odd. Such a function will be described as hermitian (see Fig. 2.6); it is often succinctly defined by the property

$$f(x) = f^*(-x),$$

and as mentioned above its Fourier transform is real. As an example of algebraic procedure for handling matters of this kind, consider that

$$f(x) = E + O + i\hat{E} + i\hat{O}.$$

Then $$f(-x) = E - O + i\hat{E} - i\hat{O}$$

and $$f^*(-x) = E - O - i\hat{E} + i\hat{O}.$$

If we now require that $f(x) = f^*(-x)$ we must have $O = 0$ and $\hat{E} = 0$. Hence $f(x) = E + i\hat{O}$.

COMPLEX CONJUGATES

The Fourier transform of the complex conjugate of a function $f(x)$ is $F^*(-s)$, that is, the *reflection* of the conjugate of the transform. Special cases of this may be summarized as follows:

$$\text{If } f(x) \text{ is } \begin{cases} \text{real} \\ \text{imaginary} \\ \text{even} \\ \text{odd} \end{cases} \text{ the transform of } f^*(x) \text{ is } \begin{cases} F(s) \\ -F(s) \\ F^*(s) \\ -F^*(s) \end{cases} = F^*(-s).$$

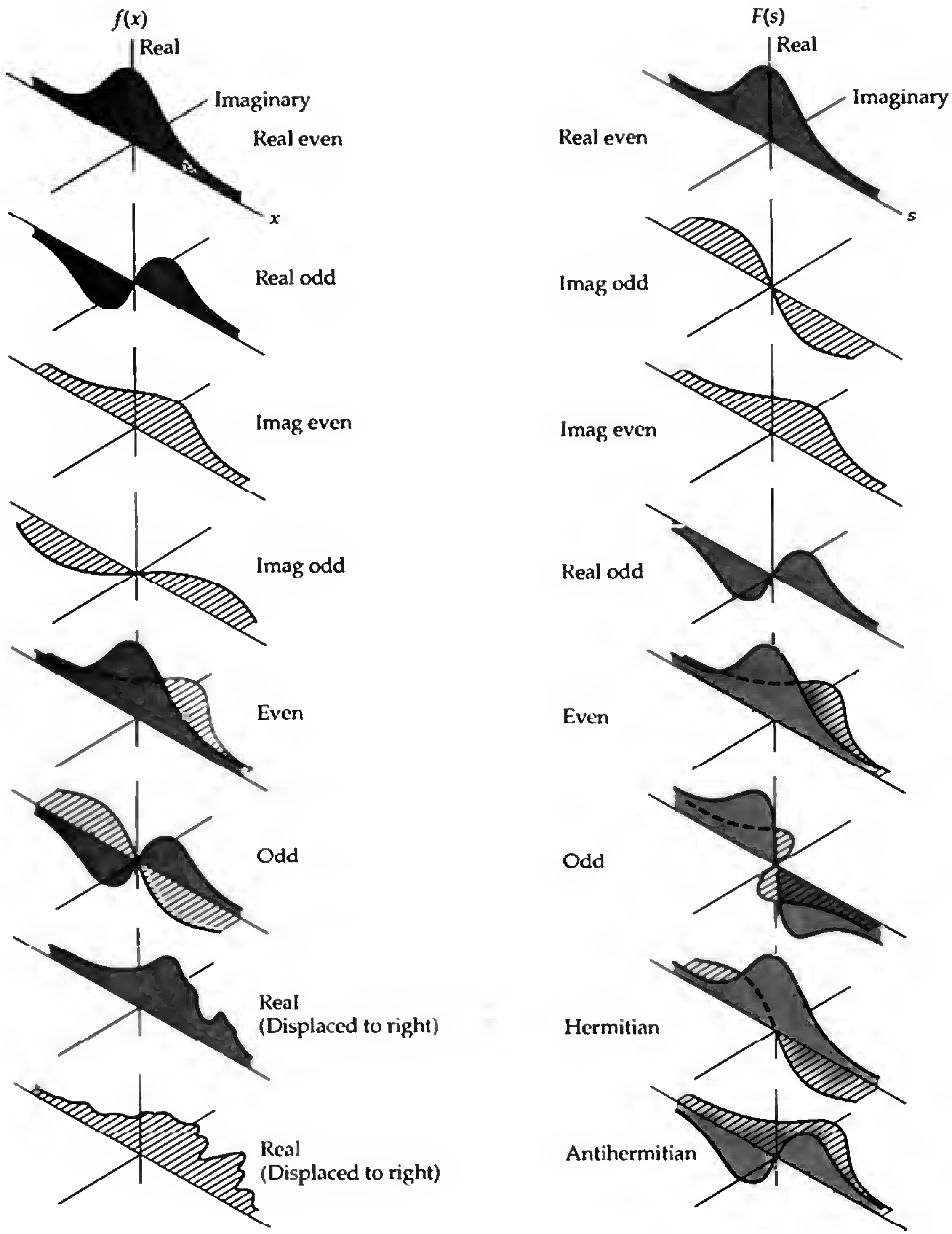


Fig. 2.5 Symmetry properties of a function and its Fourier transform.

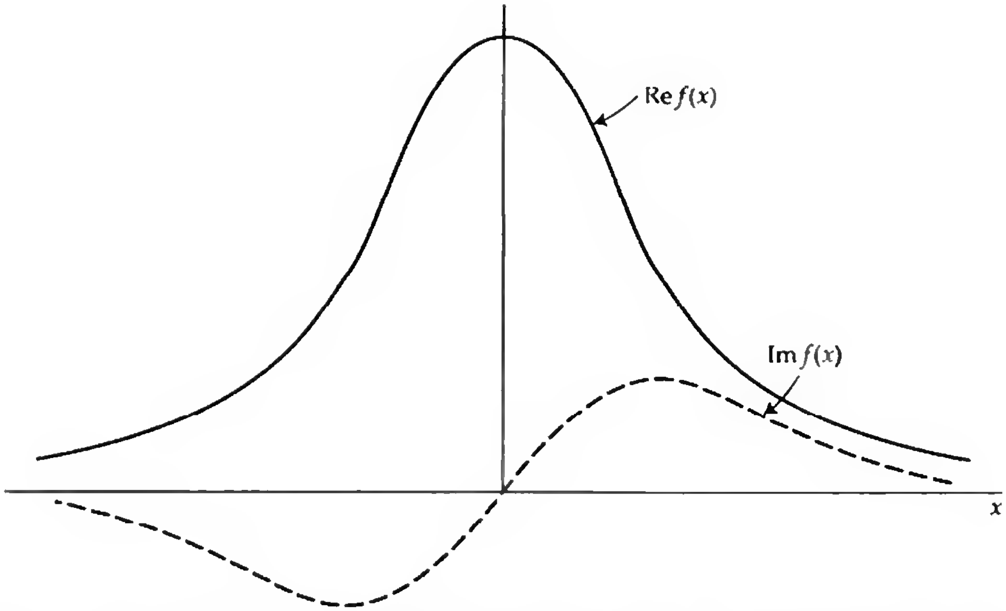


Fig. 2.6 **Hermitian functions have their real part even and their imaginary part odd. Their Fourier transform is pure real.**

Related statements are tabulated for reference.

$$
\begin{aligned}
f(x) &\supset F(s)\\
f^*(x) &\supset F^*(-s)\\
f^*(-x) &\supset F^*(s)\\
f(-x) &\supset F(-s)\\
2\,\mathrm{Re}\,f(x) &\supset F(s) + F^*(-s)\\
2\,\mathrm{Im}\,f(x) &\supset F(s) - F^*(-s)\\
f(x) + f^*(-x) &\supset 2\,\mathrm{Re}\,F(s)\\
f(x) - f^*(-x) &\supset 2\,\mathrm{Im}\,F(s)
\end{aligned}
$$

COSINE AND SINE TRANSFORMS

The cosine transform of a function $f(x)$ is defined, for positive s, as

$$2\int_0^\infty f(x)\cos 2\pi sx\,dx.$$

The cosine transform agrees with the Fourier transform if $f(x)$ is an even function. In general the even part of the Fourier transform of $f(x)$ equals the cosine transform of the even part of $f(x)$ in the region of definition, $s > 0$.

It will be noted that the cosine transform, as defined, takes no account of $f(x)$ to the left of the origin and is itself defined only to the right of the origin.

Let $F_c(s)$ represent the cosine transform of $f(x)$. Then the cosine transformation and the reverse transformation by which $f(x)$ is obtained from $F_c(s)$ are identical. Thus

$$F_c(s) = 2\int_0^\infty f(x)\cos 2\pi sx\,dx, \qquad s > 0$$

$$f(x) = 2\int_0^\infty F_c(s)\cos 2\pi sx\,ds, \qquad x > 0.$$

The sine transform of $f(x)$ is defined, for positive s, as

$$F_s(s) = 2\int_0^\infty f(x)\sin 2\pi sx\,dx.$$

This transformation is also identical with its reverse; thus

$$f(x) = 2\int_0^\infty F_s(s)\sin 2\pi sx\,dx, \qquad x > 0.$$

We may say that i times the odd part of the Fourier transform of $f(x)$ equals the sine transform of the odd part of $f(x)$, where $s > 0$.

If $f(x)$ is zero to the left of the origin, then

$$F(s) = \tfrac{1}{2}F_c(s) - \tfrac{1}{2}iF_s(s),$$

or, to restate this property for any $f(x)$,

$$\tfrac{1}{2}F_c(s) - \tfrac{1}{2}iF_s(s) = \mathcal{F} f(x)\,H(x),$$

where $H(x)$ is the unit step function (unity when x is positive and zero when x is negative).

Given $f(x)$ for $x > 0$, the cosine transform is expressible as the Fourier transform of the even function $f(x) + f(-x)$; likewise, $-i$ times the sine transform is the Fourier transform of the odd function $f(x) - f(-x)$:

$$F_c(s) = \mathcal{F}[f(x) + f(-x)] \text{ and } -iF_s(s) = \mathcal{F}[f(x) - f(-x)].$$

By looking for even (odd) functions in the Pictorial Dictionary of Fourier Transforms (Chapter 22) one readily obtains cosine (sine) transforms. Tables of cosine and sine transforms are useful for finding Fourier transforms of functions that are even or odd; major tables adopt $g(\omega) = \int_0^\infty f(x)^{\cos}_{\sin}\,\omega x\,dx$ as definitions. The apparently suppressed π raises its head when this transformation is reversed: remember that the transform of $g(x)$ is $\frac{1}{2}\pi f(\omega)$.

INTERPRETATION OF THE FORMULAS

Habitués in Fourier analysis undoubtedly are conscious of graphical interpretations of the Fourier integral. Since the integral contains a complex factor, probably the simpler cosine and sine versions are more often pictured. Thus, given $f(x)$, we picture $f(x)\cos 2\pi sx$ as an oscillation (see Fig. 2.7*a*), lying within the envelope $f(x)$ and $-f(x)$. Twice the area under $f(x)\cos 2\pi sx$ is then $F_c(s)$, for $F_c(s) = 2\int_0^\infty f(x)\cos 2\pi sx\,dx$. In Fig. 2.7*b* this area is virtually zero, but a rather high value of s is implied. Figure 2.7*c* is for a low value of s.

The Fourier integral is thus visualized for discrete values of s. The interpretation of s is important: s characterizes the frequency of the cosinusoid and is equal to *the number of cycles per unit of* x.

As an exercise in this approach to the matter, contemplate the graphical interpretation of the algebraic statements

$$F_c(s)\begin{cases} \to 0 & s\to\infty \\ = 2\int_0^\infty f(x)\,dx & s=0. \end{cases}$$

The sine transform may be pictured in the same way, and the complex transform may be pictured as a combination of even and odd parts.

A complementary and equally familiar picture results when the integrand $f(x)\cos 2\pi xs\,dx$ is regarded as a cosinusoid of amplitude $f(x)\,dx$ and frequency x; that is, is regarded as a function of s, the integral for a fixed value of x as in Fig. 2.8*a*. The same thing is shown in Fig. 2.8*b* and Fig. 2.8*c* for other discrete values of x. The summation of such curves for all values of x gives $F_c(s)$. A feeling for this approach to the transform formula is engendered in students of Fourier series by

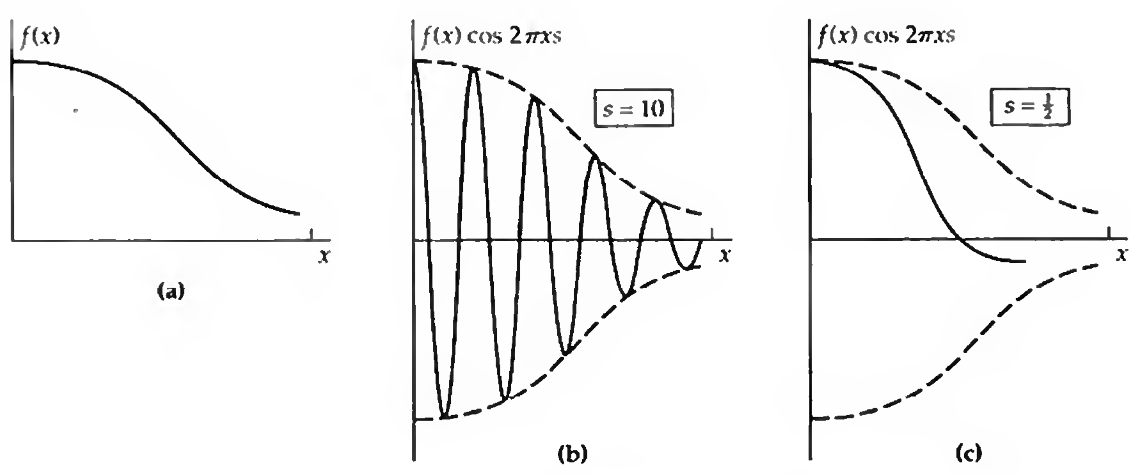


Fig. 2.7 The product of $f(x)$ with $\cos 2\pi xs$, as a function of x.

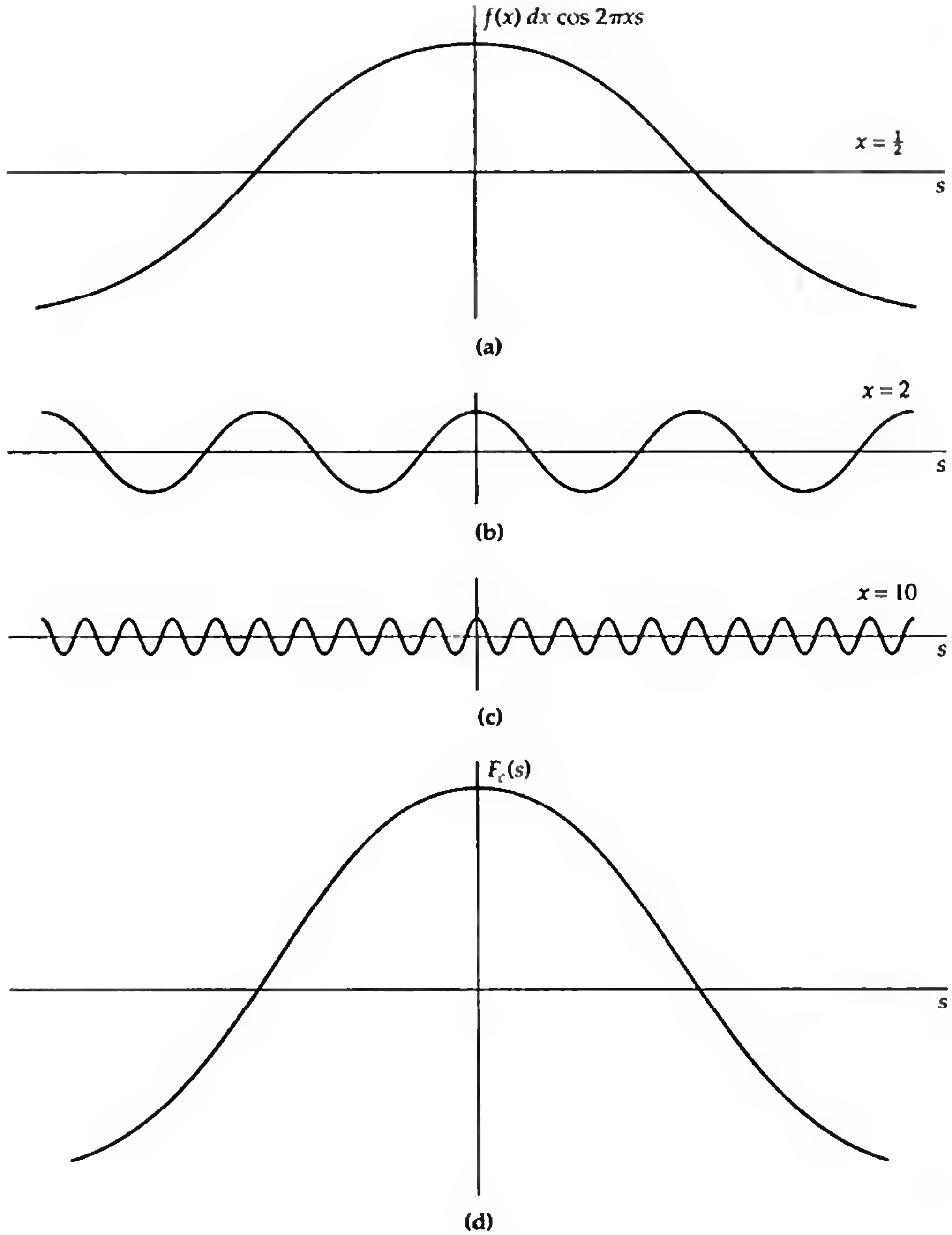


Fig. 2.8 **The product of $f(x)\,dx$ with $\cos 2\pi xs$, as a function of s.**

exercises in graphical addition of (co)sinusoids increasing arithmetically in frequency.

Each of the foregoing points of view has dual aspects, according as one ponders the analysis of a function into components or its synthesis from components. The curious fact that whether you analyze or synthesize you do the same thing

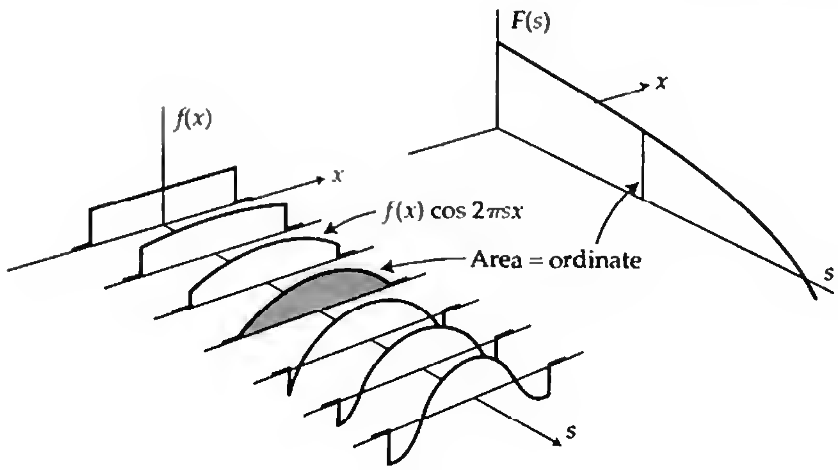


Fig. 2.9 **The surface $f(x) \cos 2\pi sx$ shown sliced in one of two possible ways.**

simply reflects the reciprocal property of the Fourier transform. Figure 2.9 illustrates the first view of the matter. If we visualize the surface represented by the slices shown for particular values of s, and then imagine it to be sliced for particular values of x, we perceive the second view.

In a further point of view we think on the complex plane (see Fig. 2.10), taking s to be fixed. The vector $f(x)\,dx$ is rotated through an angle $2\pi sx$ by the factor $\exp(-i2\pi sx)$. As $x \to \pm\infty$, the integrand $f(x)\,dx \exp(-i2\pi sx)$ shrinks in amplitude and rotates in angle, causing the integral to spiral into two limiting points, A and B. The vector AB represents the infinite integral $F(s)$ as in Fig. 2.10*a*. In Fig. 2.10*b* the more rapid coiling-up for a larger value of s is shown. The behavior of $F(s)$ as $s \to \infty$ and when $s = 0$ is readily perceived.

This kind of diagram, to which Cornu's spiral is related, is familiar from optical diffraction. It is known as a useful tool both for qualitative thinking and for numerical work in optics and antennas. It arises in the propagation of radio waves, and neatly summarizes the behavior of radio echoes reflected from ionized meteor trails as they form. Probably it would be illuminating in fields where its use is not customary.

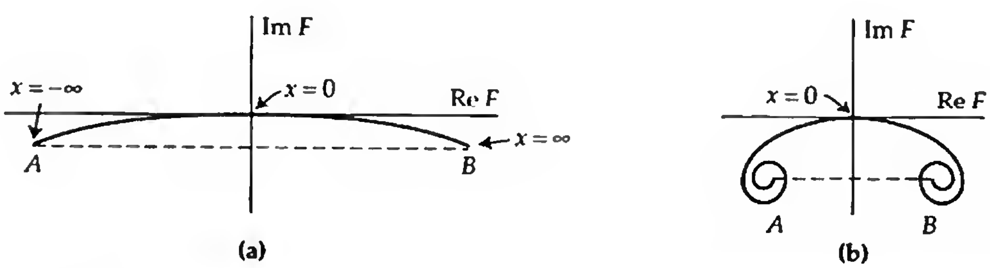


Fig. 2.10 **The Fourier integral on the complex plane of $F(s)$ when s is (a) small, (b) large.**

BIBLIOGRAPHY

Fourier integral. The following list includes standard texts on the Fourier integral and later books oriented towards electrical engineering.

Beurling, A.: "Sur les intégrales de Fourier absolument convergentes et leur application à une transformation fonctionnelle" (Neuvième congrès des mathématiciens scandinaves), Helsingfors, Finland, 1939.
Bochner, S.: "Vorlesungen über Fouriersche Integrale," Chelsea Publishing Company, New York, 1948.
Bochner, S., and K. Chandrasekharan: "Fourier Transforms," Princeton University Press, Princeton, N.J., 1949.
Bracewell, R. N.: "The Fourier Transform," *Scientific American*, vol. 260, pp. 86–95, 1989.
Bracewell, R. N.: "The Hartley Transform," Oxford University Press, New York, 1986.
Bracewell, R. N.: "Numerical Transforms," *Science*, vol. 248, pp. 697–704, 1990.
Bracewell, R. N.: "Two-Dimensional Imaging," Prentice-Hall, Englewood Cliffs, N.J., 1995.
Brigham, E. O.: "The Fast Fourier Transform," Prentice-Hall, Englewood Cliffs, N.J., 1988.
Carleman, T.: "L'Intégrale de Fourier et questions qui s'y rattachent," Almqvist and Wiskells, Uppsala, Sweden, 1944.
Carslaw, H. S.: "Introduction to the Theory of Fourier's Series and Integrals," Dover Publications, New York, 1930.
Champeney, D. C.: "Fourier Transforms and Their Physical Applications," Academic Press, London, 1973.
Gray, R. M., and J. W. Goodman: "Fourier Transforms for Engineers," Kluwer, Boston, 1995.
Körner, T. W.: "Fourier Analysis," Cambridge University Press, Cambridge, England, 1988.
Lighthill, M. J.: "An Introduction to Fourier Analysis and Generalised Functions," Cambridge University Press, Cambridge, England, 1958.
Paley, R. E. A. C., and N. Wiener: "Fourier Transforms in the Complex Domain," vol. 19, American Mathematical Society, Colloquium Publications, New York, 1934.
Papoulis, A.: "The Fourier Integral and Its Applications," McGraw-Hill Book Company, New York, 1962.
Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling: "Numerical Recipes: the Art of Scientific Programming," Cambridge University Press, Cambridge, England, 1990. See also successor volumes devoted to programs and examples in BASIC, C, FORTRAN 77, FORTAN 90, and PASCAL.
Sneddon, I. N.: "Fourier Transforms," McGraw-Hill Book Company, New York, 1951.
Titchmarsh, E. C.: "Introduction to the Theory of Fourier Integrals," Oxford University Press, Oxford, England, 1937.
Walker, J. S.: "Fourier Analysis," Oxford University Press, New York, 1988.
Wiener, N.: "The Fourier Integral and Certain of Its Applications," Cambridge University Press, Cambridge, England, 1933.

Tables of Fourier integrals. Since tables of the Laplace transform, properly interpreted, are sources of Fourier integrals, some are included here. The following are the most lengthy compilations.

Campbell, G. A., and R. M. Foster: "Fourier Integrals for Practical Applications," Van Nostrand Company, Princeton, N.J., 1948.

Doetsch, G., H. Kniess, and D. Voelker: "Tabellen zur Laplace Transformation," Springer-Verlag, Berlin, 1947.

Erdélyi, A.: "Tables of Integral Transforms," vol. 1, McGraw-Hill Book Company, New York, 1954.

McCollum, P., and B. F. Brown: "Laplace Transform Tables and Theorems," Rinehart and Wilson, New York, 1965. It has over 500 Laplace transforms of the form that arise in electric circuits.

McLachlan, N. W., and P. Humbert: "Formulaire pour le calcul symbolique," 2d ed., Gauthier-Villars, Paris, 1950.

Tables of the one-sided Laplace transformation (in which the lower limit of integration is zero) are necessarily sources of Fourier transforms only of functions that are zero for negative arguments. A table of the double-sided Laplace transform is given in the following work.

Van der Pol, B., and H. Bremmer: "Operational Calculus Based on the Two-sided Laplace Integral," 2d ed., Cambridge University Press, Cambridge, England, 1955.

PROBLEMS

1. What condition must $F(s)$ satisfy in order that $f(x) \to 0$ as $x \to \pm\infty$?

2. Prove that $|F(s)|^2$ is an even function if $f(x)$ is real.

3. The Fourier transform in the limit of sgn x is $(i\pi s)^{-1}$. What conditions for the existence of Fourier transforms are violated by these two functions? (sgn x equals 1 when x is positive and -1 when x is negative.)

4. Show that all periodic functions violate a condition for the existence of a Fourier transform.

5. Verify that the function cos x violates one of the conditions for existence of a Fourier transform. Prove that $\exp(-\alpha x^2)\cos x$ meets this condition for any positive value of α.

6. Give the odd and even parts of $H(x)$, e^{ix}, $e^{-x}H(x)$, where $H(x)$ is unity for positive x and zero for negative x.

7. Graph the odd and even parts of $[1 + (x - 1)^2]^{-1}$.

8. Show that the even part of the product of two functions is equal to the product of the odd parts plus the product of the even parts.

9. Investigate the relationship of $\mathcal{F}\mathcal{F}f$ to f when f is neither even nor odd.

10. Show that $\mathcal{F}\mathcal{F}\mathcal{F}\mathcal{F}f = f$.

11. It is asserted that the odd part of log x is a constant. Could this be correct?

12. Is an odd function of an odd function an odd function? What can be said about odd functions of even functions and even functions of odd functions?

13. Prove that the Fourier transform of a real odd function is imaginary and odd. Does it matter whether the transform is the plus-i or minus-i type?

14. An antihermitian function is one for which $f(x) = -f^*(-x)$. Prove that its real part is odd and its imaginary part even and thus that its Fourier transform is imaginary. ▷

15. Point out the fallacy in the following reasoning. "Let $f(x)$ be an odd function. Then the value of $f(-a)$ must be $-f(a)$; but this is not the same as $f(a)$. Therefore an odd function cannot be even."

16. Let the odd and even parts of a function $f(x)$ be $o(x)$ and $e(x)$. Show that, irrespective of shifts of the origin of x,

$$\int_{-\infty}^{\infty} |o(x)|^2\,dx + \int_{-\infty}^{\infty} |e(x)|^2\,dx = \text{const.}$$

17. Note that the odd and even parts into which a function is analyzed depend upon the choice of the origin of abscissas. Yet the sum of the integrals of the squares of the odd and even parts is a constant that is independent of the choice of origin. What is the constant?

18. Let axes of symmetry of a real function $f(x)$ be defined by values of a such that if o and e are the odd and even parts of $f(x - a)$, then

$$\frac{\int o^2\,dx - \int e^2\,dx}{\int o^2\,dx + \int e^2\,dx}$$

has a maximum or minimum with respect to variation of a. Show that all functions have at least one axis of symmetry. If there is more than one axis of symmetry, can there be arbitrary numbers of each of the two kinds of axis?

19. Note that $\cos x$ is fully even and has no odd part, and that shift of origin causes the even part to diminish and the odd part to grow until in due course the function becomes fully odd. In fact the even part of any periodic function will wax and wane relative to the odd part as the origin shifts. Consider means of assigning "abscissas of symmetry" and quantitative measures of "degree of symmetry" that will be independent of the origin of x. Test the reasonableness of your conclusions—for example, on the functions of period 2 which in the range $-1 < x < 1$ are given by $\Lambda(x)$, $\Lambda(x) - \frac{1}{2}$, $\Lambda(x) - \frac{1}{4}$. See Chapter 4 for triangle-function notation $\Lambda(x)$. ▷

20. The function $f(x)$ is equal to unity when x lies between $-\frac{1}{2}$ and $\frac{1}{2}$ and is zero outside. Draw accurate loci on the complex plane of $F(s)$ from which values of $F(0)$, $F(\frac{1}{2})$, $F(1)$, $F(1\frac{1}{2})$, and $F(2)$ can be measured.

21. The function $f(x)$ is equal to 100 when x differs by less than 0.01 from 1, 2, 3, 4, or 5 and is zero elsewhere. Draw a locus on the complex plane of $F(s)$ from which $F(0.05)$ can be measured in amplitude and phase.

22. **Self-transforming functions.** Long-published tables of Fourier transforms have included two functions that transform into themselves, namely $\exp(-\pi x^2)$ and sech x. In 1956, III(x) joined this select group. When impulses are included, other examples, such as $1 + \delta(x)$, can be given. Propose a general construction for self-transforming functions. ▷

3

Convolution

The *idea* of the convolution of two functions occurs widely, as witnessed by the multiplicity of its aliases. The *word* "convolution" is coming into more general use as awareness of its oneness spreads into various branches of science. The German term *Faltung* is occasionally seen, as is the term "composition product," adapted from the French. Terms encountered in special fields include superposition integral, Duhamel integral, Borel's theorem, (weighted) running mean, cross-correlation function, smoothing, blurring, scanning, and smearing.

As some of these last terms indicate, convolution describes the action of an observing instrument when it takes a weighted mean of some physical quantity over a narrow range of some variable. When, as very often happens, the form of the weighting function does not change appreciably as the central value of the variable changes, the observed quantity is a value of the convolution of the distribution of the desired quantity with the weighting function, rather than a value of the desired quantity itself. All physical observations are limited in this way by the resolving power of instruments, and for this reason alone convolution is ubiquitous. Later we show that the appearance of convolution is coterminous with linearity plus time or space invariance, and also with sinusoidal response to sinusoidal stimulus.

Not only is convolution widely significant as a physical concept, but because of a powerful theorem encountered below, it also offers an advantageous starting point for theoretical developments. Conversely, because of its adaptability to computing, it is an advantageous terminal point for theory before one starts numerical work.

The convolution of two functions $f(x)$ and $g(x)$ is

$$\int_{-\infty}^{\infty} f(u)g(x-u)\,du,$$

or briefly,

$$f(x) * g(x).$$

The convolution itself is also a function of x, let us say $h(x)$.

Various ways of looking at the convolution integral suggest themselves. For example, suppose that $g(x)$ is given. Then for every function $f(x)$ for which the integral exists there will be an $h(x)$. Following Volterra, we may say that $h(x)$ is a *functional* of the function $f(x)$. Note that to calculate $h(x_1)$ we need to know $f(x)$ for a whole range of x, whereas to calculate a *function* of the function $f(x)$ at $x = x_1$, we need only know $f(x_1)$.

In Fig. 3.1 the product $f(u)g(x - u)$ is shown shaded, and the ordinate $h(x)$ is equal to the shaded area.

A second example with a different $f(x)$ but the same $g(x)$ is shown in Fig. 3.2 to illustrate the general features relating the functional $h(x)$ to $f(x)$. It will be seen that $h(x)$ is smoother in detail than $f(x)$, is more spread out, and has less total variation.

In another approach, $f(x)$ is resolved into infinitesimal columns (see Fig. 3.3). Each column is regarded as melted out into heaps having the form of $g(x)$ but centered at its original value of x. Just two of these melted-down columns are shown in the figure, but all are to be so pictured. Then $h(x)$ is equal to the sum of the contributions at the point x made by all the heaps; that is,

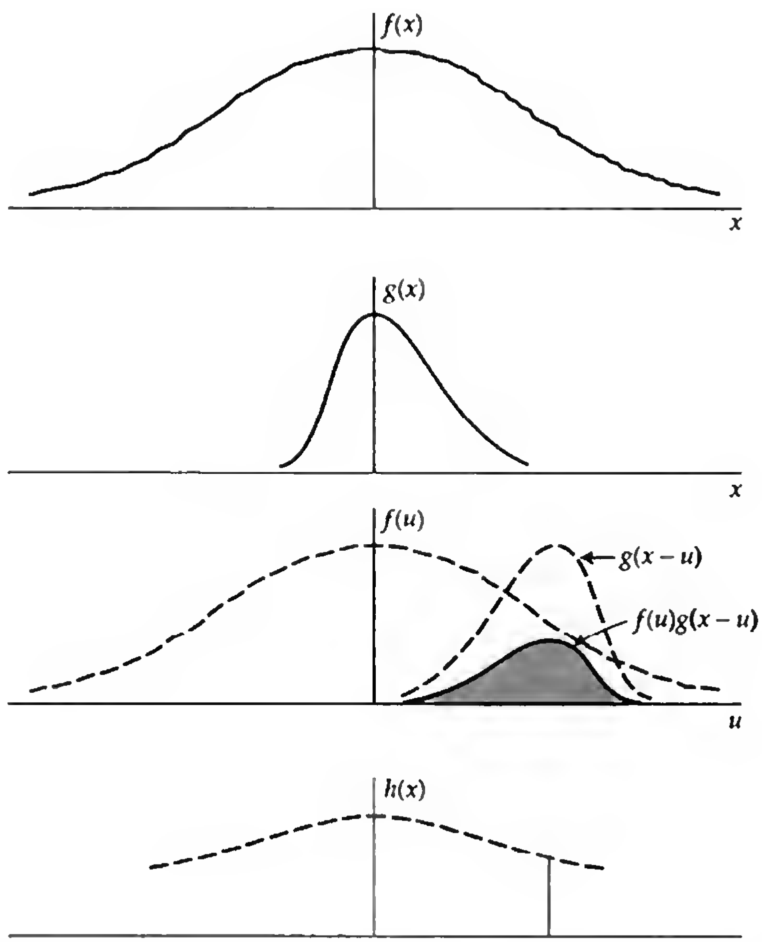


Fig. 3.1 **The convolution integral $h(x) = f(x) * g(x)$ represented by a shaded area.**

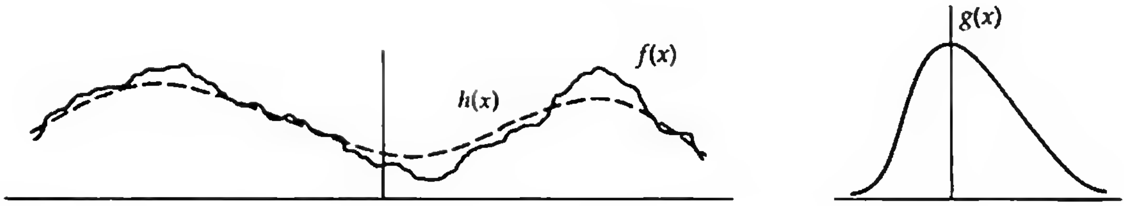


Fig. 3.2 **Illustrating the smoothing effect of convolution ($h = f * g$).**

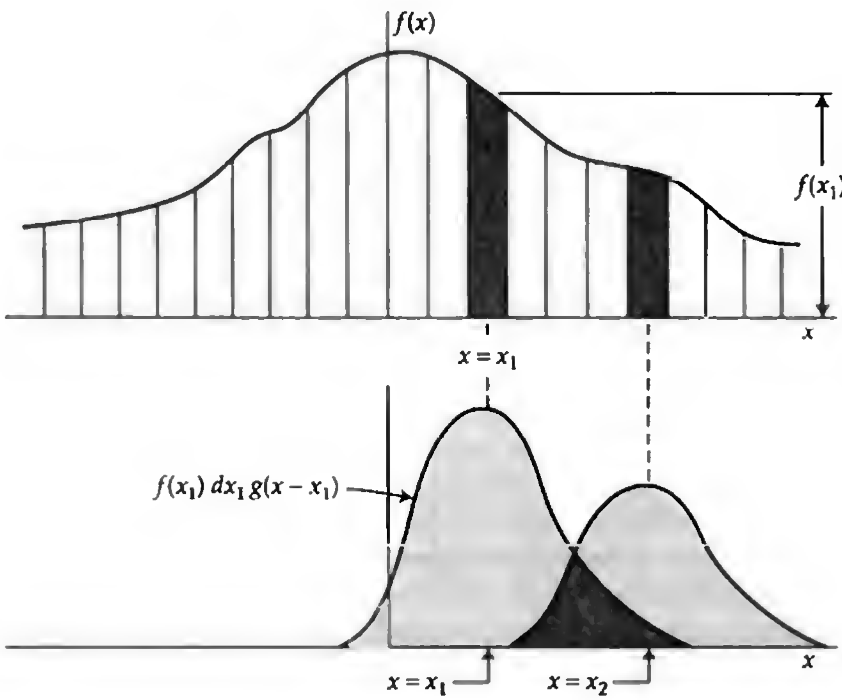


Fig. 3.3 **The convolution integral regarded as a superposition of characteristic contributions.**

$$h(x) = \int_{-\infty}^{\infty} f(x_1)g(x - x_1)\,dx_1.$$

A further view is illustrated in Fig. 3.4, where $g(u)$ is shown folded back on itself about the line $u = \frac{1}{2}x$. As before, the area under the product curve $f(u)g(x - u)$ is the convolution $h(x)$, and its dependence on the position of the line of folding (German *Faltung*) may often be visualized from this standpoint.

These suggestions do not exhaust the possible interpretations of the convolution integral; others will appear later.

It should be noticed that before the operations of multiplication and integration take place $g(x)$ must be reversed. It is possible to dispense with the reversal, and in some subjects this is more natural. A consequence of the reversal is, however, that convolution is commutative; that is,

$$f * g = g * f,$$

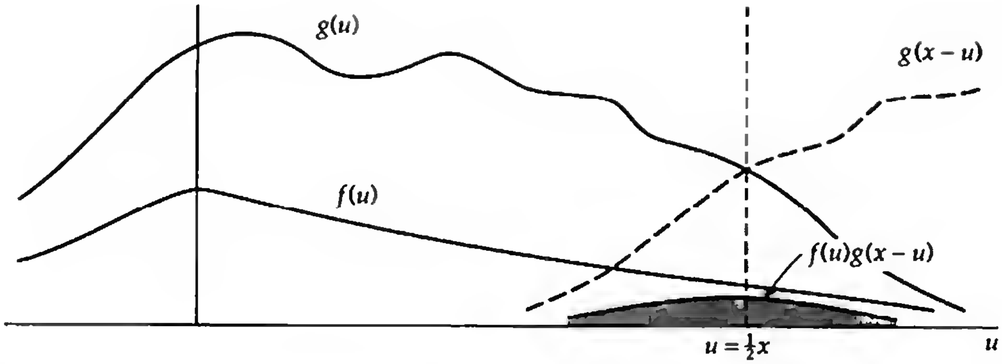


Fig. 3.4 **Convolution from the *Faltung* standpoint.**

or
$$\int_{-\infty}^{\infty} f(u)g(x-u)\,du = \int_{-\infty}^{\infty} g(u)f(x-u)\,du.$$

Convolution is also associative (provided that all the convolution integrals exist),

$$f * (g * h) = (f * g) * h,$$

and distributive over addition,

$$f * (g + h) = f * g + f * h.$$

The abbreviated notation with asterisks (*) thus proves very convenient in formal manipulation, since the asterisks behave like multiplication signs.

EXAMPLES OF CONVOLUTION

Consider the truncated exponential function

$$E(x) = \begin{cases} e^{-x} & x > 0 \\ 0 & x < 0. \end{cases}$$

We shall calculate the convolution between two such truncated exponentials with different (positive) decay constants. Thus

$$\begin{aligned} aE(\alpha x) * bE(\beta x) &= ab \int_{-\infty}^{\infty} E(\alpha u)\, E(\beta x - \beta u)\, du \\ &= abE(\beta x) \int_0^x E(\alpha u - \beta u)\, du \\ &= abE(\beta x) \frac{E(\alpha x - \beta x) - 1}{\beta - \alpha} \\ &= ab \frac{E(\alpha x) - E(\beta x)}{\beta - \alpha}. \end{aligned}$$

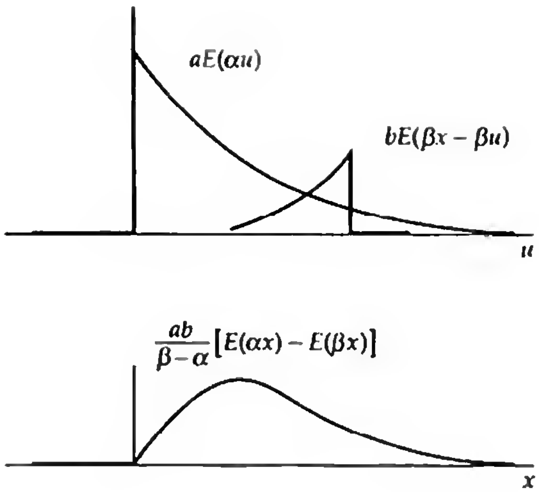


Fig. 3.5 **Convolution of two truncated exponentials.**

The result is thus the difference of two truncated exponentials, each with the same amplitude, as illustrated in Fig. 3.5. This function occurs commonly; for instance, it describes the concentration of a radioactive isotope which decays with a constant α while simultaneously being replenished as the decay product of a parent isotope which decays with a constant β. From the commutative property of convolution we see that the result is the same if α and β are interchanged, and as $t \to \infty$ one of the terms dies out, leaving a simple exponential with constant α or β, whichever describes the slower decay. As a special case we calculate $E(\alpha x) * E(\alpha x)$ by taking the limit as $\beta - \alpha \to 0$. Thus we find

$$\begin{aligned} E(\alpha x) * E(\alpha x) &= \lim_{\beta-\alpha\to 0} \frac{E(\alpha x) - E(\beta x)}{\beta - \alpha} \\ &= -\frac{d}{d\alpha} E(\alpha x) \\ &= xE(\alpha x). \end{aligned}$$

This particular function describes the response of a critically damped resonator, such as a dead-beat galvanometer, to an impulsive disturbance.

As a further example consider $E(-\alpha x) * E(\beta x)$. This gives an entirely different type of result. Thus

$$\begin{aligned} E(-\alpha x) * E(\beta x) &= \int_{-\infty}^{\infty} E(-\alpha x + \alpha u)E(\beta u)\,du \\ &= \begin{cases} \displaystyle\int_x^{\infty} e^{-\alpha u + \alpha x} e^{-\beta u}\,du & x > 0 \\ \displaystyle\int_0^{\infty} e^{-\alpha u + \alpha x} e^{-\beta u}\,du & x < 0 \end{cases} \\ &= \frac{E(-\alpha x) + E(\beta x)}{\alpha + \beta}. \end{aligned}$$

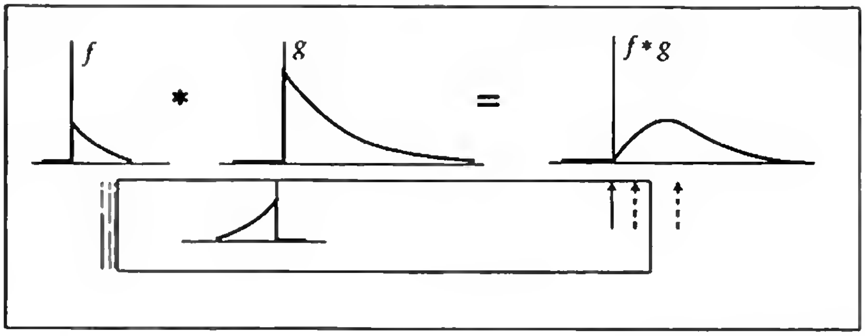


Fig. 3.6 **Graphical construction for convolution. The movable piece of paper has a graph of one of the functions plotted backward.**

Here we have a function that is peaked at the origin and dies away with a constant α to the left and with a constant β to the right.

In calculations of this kind care is required in fixing the limits of the integrals and checking signs, because the ordinary sort of algebraic error can make a radical change in the result. The following graphical construction for convolution is useful as a check. Plot one of the functions entering into the convolution backward on a movable piece of paper as shown in Fig. 3.6, and slide it along in the direction of the axis of abscissas. When the movable piece is to the left of the position shown, the product of g with f reversed is zero. By marking an arrow in some convenient position, we can keep track of this. Then suddenly, at the position shown, the integral of the product begins to assume nonzero values. By moving the paper a little farther along, as indicated by a broken outline, we find that the convolution will be positive and increasing from zero approximately linearly with displacement. Farther along still, we see that a maximum will occur beyond which the convolution dies away.

In the second example given above, two oppositely directed exponential tails, the absence of a zero stretch, and the presence of a jump in slope at the origin are immediately apparent from the construction. In addition to qualitative conclusions such as these, certain quantitative results are yielded by this construction in

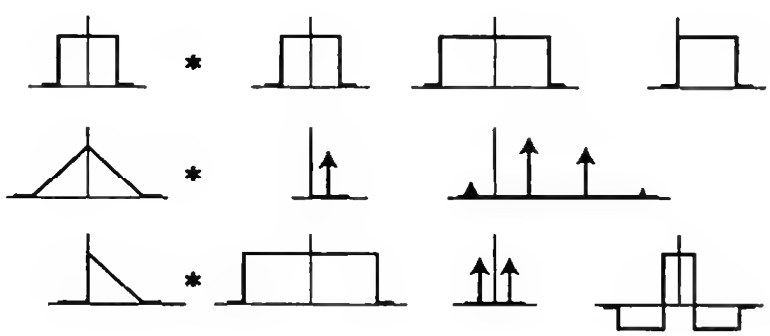


Fig. 3.7 **Examples for practicing graphical convolution. The arrows represent impulses.**

many cases, particularly where the functions break naturally into parts in successive ranges of the abscissa.

It is well worth while to carry out this moving construction until it is thoroughly familiar. There is no doubt that experts do this geometrical construction in their heads all the time, and a little practice soon enables the beginner also to dispense with the actual piece of paper. Examples for practice are given in Fig. 3.7.

SERIAL PRODUCTS

Consider two polynomials

$$a_0 + a_1x + a_2x^2 + a_3x^3 + \dots$$

and

$$b_0 + b_1x + b_2x^2 + b_3x^3 + \dots .$$

Their product is

$$a_0b_0 + (a_0b_1 + a_1b_0)x + (a_0b_2 + a_1b_1 + a_2b_0)x^2 + (a_0b_3 + a_1b_2 + a_2b_1 + a_3b_0)x^3 + \dots ,$$

which we may call

$$c_0 + c_1x + c_2x^2 + c_3x^3 + \dots ,$$

where
$$\begin{aligned} c_0 &= a_0b_0 \\ c_1 &= a_0b_1 + a_1b_0 \\ c_2 &= a_0b_2 + a_1b_2 + a_2b_0 \\ c_3 &= a_0b_3 + a_1b_2 + a_2b_1 + a_3b_0. \end{aligned}$$

This elementary observation has an important connection with convolution. Suppose that two functions f and g are given, and that it is required to calculate their convolution numerically. We form a sequence of values of f at short regular intervals of width w,

$$\{f_0\ f_1\ f_2\ f_3 \cdots f_m\},$$

and a corresponding sequence of values of g,

$$\{g_0\ g_1\ g_2\ g_3\ \cdots\ g_n\}.$$

We then approximate the convolution integral

$$\int_{-\infty}^{\infty} f(x')g(x - x')\,dx'$$

by summing products of corresponding values of f and g, taking different discrete values of x one by one. It is convenient to write the g sequence on a movable strip of paper which can be slid into successive positions relative to the f sequence for each successive value. The first few stages of this approach are shown in Fig. 3.8. It will be noticed that the g sequence has been written in reverse as required by

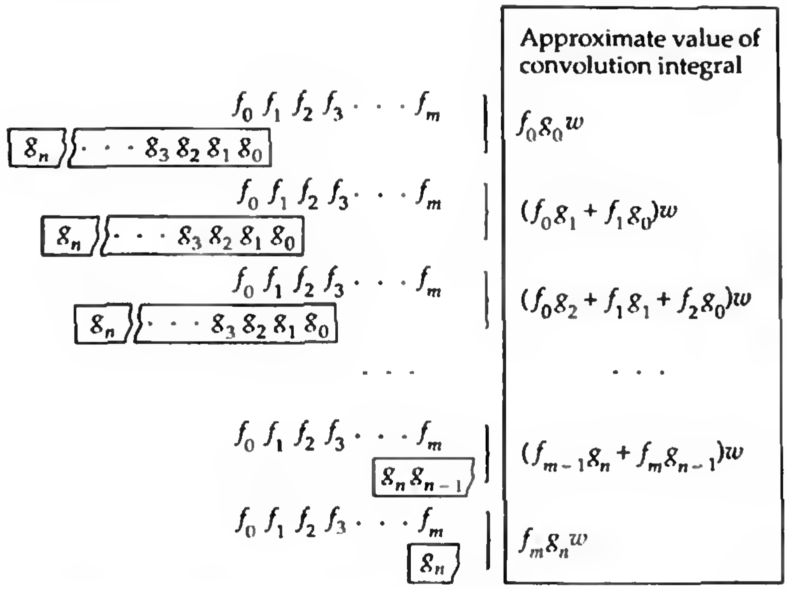


Fig. 3.8 Explaining the serial product.

the formula. Since $f * g = g * f$, the f sequence could have been written in reverse, in which case *it* would have been written on the movable strip.

It will be seen that this procedure generates the same expressions that occur in the multiplication of series, and we therefore introduce the term "serial product" to describe the sequence of numbers

$$\{f_0 g_0 \quad f_0 g_1 + f_1 g_0 \quad f_0 g_2 + f_1 g_1 + f_2 g_0 \quad \cdots\}$$

derived from the two sequences

$$\{f_0 \;\; f_1 \;\; f_2 \;\; f_3 \cdots\} \qquad \text{and} \qquad \{g_0 \;\; g_1 \;\; g_2 \;\; g_3 \cdots\}.$$

We transfer the asterisk notation to represent this relationship between the three sequences as follows:

$$\{f_0 \, f_1 \cdots f_m\} * \{g_0 g_1 \cdots g_n\} = \{f_0 g_0 \;\; f_0 g_1 + f_1 g_0 \;\; \cdots f_m g_n\}.$$

Alternatively, we may define the $(i + 1)$th term of the serial product of $\{f_i\}$ and $\{g_i\}$ to be

$$\sum_j f_j g_{i-j}.$$

In practice the calculation of serial products is an entirely feasible procedure (for example, on a hand calculator by allowing successive products to accumulate in the product register). The two sequences are most conveniently written in

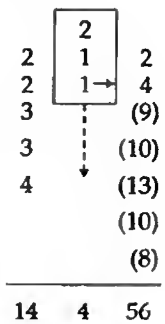


Fig. 3.9 Calculating a serial product by hand.

vertical columns, and the answers are written opposite an arrow marked in a convenient place on the movable strip. Figure 3.9 shows an early stage in the calculation of $\{2\,2\,3\,3\,4\} * \{1\,1\,2\}$. The value of 4, shown opposite the arrow, has just been calculated, and the values still to be calculated as the movable strip is taken downward are shown in parentheses. Note that the sequence on the *moving* strip has been written in *reverse* (upward).

It will be seen that the serial product is a longer sequence than either of the component sequences, the number of terms being one less than the sum of the numbers of terms in the components:

$$\{2\,2\,3\,3\,4\} * \{1\,1\,2\} = \{2\,4\,9\,10\,13\,10\,8\}.$$

Furthermore the sum of the terms of the serial product is the product of the sums of the component sequences, a fact which allows a very valuable check on numerical work. As a special case, if the sum of one of the component sequences is unity, then the sum of the serial product will be the same as the sum of the other component. These properties are analogous to properties of the convolution integral.

A semi-infinite sequence is one such as

$$\{f_0\ f_1\ f_2 \cdots\},$$

which has an end member in one direction but runs on without end in the other. If we take the serial product of such a sequence with a finite sequence, the result is also semi-infinite; for example,

$$\{1\,1\} * \{1\,2\,3\,4\,\ldots\} = \{1\,3\,5\,7\,9\,\ldots\}.$$

A semi-infinite sequence may or may not have a finite sum, but this does not lead to problems in defining the serial product because each member is the sum of only a finite number of terms. Of course, the numerical check, according to which the sum of the members of the serial product equals the product of the sums, breaks down if any one of the sums does not exist.

The serial product of two semi-infinite sequences presents no problems when both run on indefinitely in the same direction; thus

$$\{1\,2\,3\,4\,\ldots\} * \{1\,2\,3\,4\,\ldots\} = \{1\,4\,10\,20\,\ldots\},$$

but if they run on in opposite directions then each member of the serial product is the sum of an infinite number of terms. This sum may very well not exist; thus

$$\{\dots 4\,3\,2\,1\} * \{1\,2\,3\,4\dots\}$$

does not lead to convergent series.

Two-sided sequences are often convenient to deal with and call for no special comment other than that one must specify the origin explicitly, for example, by an arrow as in

$$\{\dots 0.1\ \ 0.2\ \ 0.4\ \ \underset{\uparrow}{0.9}\ \ 0.8\ \ 0.7\ \ 0.6\dots\}.$$

The sequence

$$\{\dots 0\,0\,0\,\underset{\uparrow}{1}\,0\,0\,0\dots\} = \{I\}$$

plays an important role analogous to that of the impulse symbol $\delta(x)$. It has the property that

$$\{I\} * \{f\} = \{f\}$$

for all sequences $\{f\}$. This property is, of course, also possessed by the one-member sequence $\{1\}$ and by other sequences such as $\{1\ 0\ 0\}$.

Serial multiplication by the sequence

$$\{\dots 0\,0\,0\,1\ -1\,0\,0\,0\dots\} \qquad \text{or} \qquad \{1\ -1\}$$

is equivalent to taking the first finite difference; that is,

$$\begin{aligned}\{1\ -1\} * \{\dots f_{-2}\,f_{-1}\,f_0\,f_1\,f_2\dots\}\\ = \{\dots\ \ f_{-1}-f_{-2}\quad f_0-f_{-1}\quad f_1-f_0\quad f_2-f_1\ \ \dots\}.\end{aligned}$$

Where it is important to distinguish between the central difference $f_{n+\frac{1}{2}} - f_{n-\frac{1}{2}}$ and the forward difference $f_{n+1} - f_n$ it is necessary to specify the origins of the sequences appropriately. Apart from a shift in the indexing of the terms, however, the resulting sequence is the same.

The sequence

$$\left\{\frac{1}{n}\ \frac{1}{n}\ \dots (n \text{ terms}) \dots \frac{1}{n}\right\}$$

generates the running mean over n terms, a familiar process for smoothing sequences of meteorological data. It may be written, as in the case of weekly running means of daily values, as

$$\frac{1}{7}\{1\,1\,1\,1\,1\,1\,1\}.$$

The sequence $\{1\ 1\ 1\ 1\ 1\ 1\ 1\}$ itself generates running sums, in this case over seven terms.

The semi-infinite sequence

$$\{1\,1\,1\,1\,\ldots\}$$

enters into a serial product with another sequence to generate a familiar result. Thus if $\{f\}$ is a sequence and S_n is the sum of the first n terms, then the sequence $\{S\}$ defined by

$$\{S\} = \{S_1\,S_2\,S_3\,\ldots\}$$

may be expressed in the form

$$\{1\,1\,1\,1\,\ldots\} * \{f\} = \{S\}.$$

Since each term of the sequence $\{f\}$ is the difference between two successive terms of $\{S\}$, a converse relationship can also be written:

$$\{1\ {-1}\} * \{S\} = \{f\}.$$

Substituting in the previous equation, we have

$$\{1\,1\,1\,1\,\ldots\} * \{1\ {-1}\} * \{S\} = \{S\};$$

in other words the two operations neutralize each other. We may express this reciprocal relationship by writing

$$\{1\,1\,1\,1\,\ldots\} * \{1\ {-1}\} = \{1\,0\,0\,0\,\ldots\}$$

or, in a two-sided version,

$$\{\ldots\,0\,0\,0\,\underset{\uparrow}{1}\,1\,1\,1\,\ldots\} * \{\ldots\,0\,0\,0\,\underset{\uparrow}{1}\,{-1}\,0\,0\,\ldots\} = \{\ldots\,0\,0\,0\,1\,0\,0\,0\,\ldots\}.$$

Another example of a reciprocal pair of sequences is

$$\{1\,2\,3\,4\,\ldots\} \qquad \text{and} \qquad \{1\ {-2}\ 1\}.$$

This is an important relationship because the reciprocal sequence represents the solution to the problem of finding $\{g\}$ when $\{f\}$ and $\{h\}$ are given and it is known that

$$\{f\} * \{g\} = \{h\}.$$

The answer is

$$\{g\} = \{f\}^{-1} * \{h\},$$

where $\{f\}^{-1}$ is the reciprocal of $\{f\}$, that is, the sequence with the property that

$$\{f\}^{-1} * \{f\} = \{\ldots\,0\,0\,1\,0\,0\,\ldots\}.$$

Inversion of serial multiplication. If $\{f\} * \{g\} = \{h\}$, then $\{h\}$ is called the serial product of $\{f\}$ and $\{g\}$, because the sequence $\{h\}$ comprises the coefficients of the polynomial which is the product of the polynomials represented by $\{f\}$ and $\{g\}$.

Conversely, the process of finding $\{g\}$ when $\{f\}$ and $\{h\}$ are given may be called serial division, and in fact one could carry out the solution of such a problem by actually doing long division of polynomials.

For example, to solve the problem

$$\{1\ 1\} * \{g_0\, g_1\, g_2 \cdots\} = \{1\ 3\ 3\ 1\}$$

one could write

$$g_0 + g_1 x + g_2 x^2 + \ldots = \frac{1 + 3x + 3x^2 + x^3}{1 + x}.$$

The long division would then proceed as follows:

$$\begin{array}{r l}
 & \quad 1 + 2x + x^2 \\
1 + x\,) & \overline{1 + 3x + 3x^2 + x^3} \\
 & \underline{1 + x} \\
 & \qquad 2x + 3x^2 \\
 & \qquad \underline{2x + 2x^2} \\
 & \qquad\qquad\quad x^2 + x^3 \\
 & \qquad\qquad\quad \underline{x^2 + x^3}
\end{array}$$

Thus
$$g_0 + g_1 x + g_2 x^2 + \ldots = 1 + 2x + x^2$$
or
$$\{g_0\, g_1\, g_2 \cdots\} = \{1\ 2\ 1\}.$$

This, of course, is recognizable immediately as the correct answer.

However, it is not necessary to organize the calculation at such length. A convenient way of doing the job on a hand calculator may be shown by going back to an example in the previous section, where $\{2\ 2\ 3\ 3\ 4\} * \{1\ 1\ 2\}$ was calculated. The situation at the end of the calculation is as shown in Fig. 3.10*a*, the serial product being in the right-hand column. Now, suppose that the left-hand column $\{f\}$

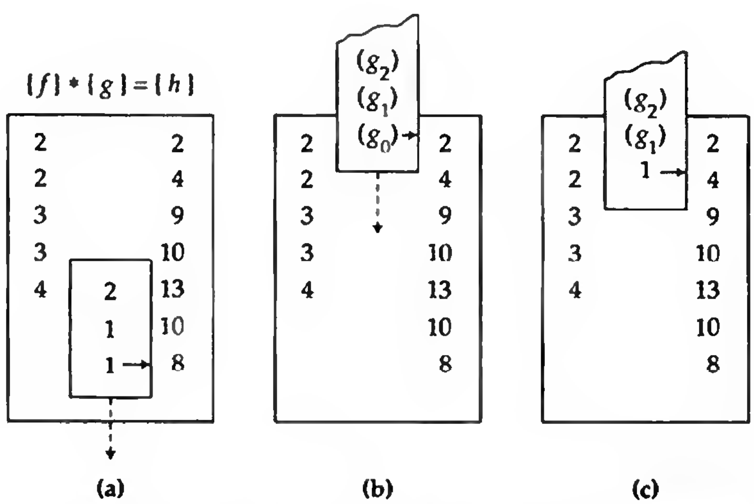


Fig. 3.10 **(a) Completion of serial multiplication of {2 2 3 3 4} by{1 1 2}, (b) and (c) first steps in the inversion of the process.**

and the right-hand column $\{h\}$ were given. To find $\{g\}$, write the sequence upward on the movable strip of paper, and place it in position for beginning the calculation of the serial product (see Fig. 3.10*b*). Clearly g_0 is immediately deducible, for it has to be such that when it is multiplied by the 2 on the left the result is 2. Hence in the space labeled g_0 one may write 1 and then move the paper down to the next position (see Fig. 3.10*c*). Then, when g_1 is multiplied by 2 and added to the product of 2 and 1, the result must be 4. This gives g_1, and so on.

With a hand calculator one forms the products of the numbers in the left-hand column with such numbers on the moving strip as are already known and accumulates the sum of the products. Subtract this sum from the current value of $\{h\}$ opposite the arrow and divide by the first member of $\{f\}$ at the top of the left column to obtain the next value of $\{g\}$. Enter this value on the paper strip, which may then be moved down one more place. The process is started by entering the first value of $\{g\}$, which is h_0/f_0. This inversion of serial multiplication is as easy as the direct process and takes very little longer.

Compared with long division, this method is superior, since it involves writing down no numbers other than the data and the answer. Furthermore it is as readily applicable when the coefficients are large numbers or fractions. Long division might, however, be considered for short sequences of small integers.

Table 3.1 lists a few common sequences and their inverses. Most of the entries may be verified by serial multiplication of corresponding entries, whereupon the result will be found to be $\{1\ 0\ 0\ 0\ \ldots\}$. The table is precisely equivalent to a list of polynomial pairs beginning as follows:

$$\begin{array}{ll}
1 + x & 1 - x + x^2 - x^3 + \ldots \\
1 - x & 1 + x + x^2 + x^3 + \ldots \\
1 + x^2 & 1 - x^2 + x^4 + \ldots \\
\ldots & \ldots
\end{array}$$

In this case the property of corresponding entries would be that their product was unity.

Some of the later entries in Table 3.1 represent exponential and other variations reminiscent of the natural behavior of circuits, and will prove more difficult, though not impossible, to verify at this stage. They are readily derivable by transform methods.

The serial product in matrix notation. Let the sequence $\{h\}$ be the serial product of two sequences $\{f\}$ and $\{g\}$,

$$\begin{aligned}
\text{where } \{f\} &= \{f_0\, f_1\, f_2 \cdots f_m\} \\
\{g\} &= \{g_0\, g_1\, g_2 \cdots g_n\} \\
\{h\} &= \{h_0\, h_1\, h_2 \ldots h_{m+n}\}.
\end{aligned}$$

Evidently all three sequences can be expressed as single-row or single-column matrices, but since the sequences have different numbers of members, matrix notation might not immediately spring to mind. The relationship between $\{f\}$, $\{g\}$, and $\{h\}$ can, however, be expressed in terms of matrix multiplication as follows.

TABLE 3.1
Some sequences and their inverses

Sequence	Inverse
$\{1\ 1\}$	$\{1\ -1\ 1\ -1\ 1\ -1\ \ldots\}$
$\{1\ -1\}$	$\{1\ 1\ 1\ 1\ 1\ 1\ \ldots\}$
$\{1\ 0\ 1\}$	$\{1\ 0\ -1\ 0\ 1\ 0\ \ldots\}$
$\{1\ 0\ -1\}$	$\{1\ 0\ 1\ 0\ 1\ 0\ \ldots\}$
$\{1\ 1\ 1\}$	$\{1\ -1\ 0\ 1\ -1\ 0\ \ldots\}$
$\{1\ 2\ 1\}$	$\{1\ -2\ 3\ -4\ 5\ -6\ \ldots\}$
$\{n+1\} = \{1\ 2\ 3\ 4\ \ldots\}$	$\{1\ -1\}^{*2} = \{1\ -2\ 1\}$
$\{(n+1)^2\} = \{1\ 4\ 9\ 16\ \ldots\}$	$\{1\ -1\}^{*3} * \{1\ 1\}^{-1}$
$\{(n+1)^3\} = \{1\ 8\ 27\ \ldots\}$	$\{1\ -1\}^{*4} * \{1\ 4\ 1\}^{-1}$
$\{1\ -a\}$	$\{1\ a\ a^2\ a^3\ \ldots\}$
$\{1\ -a\}^{*2}$	$\{1\ 2a\ 3a^2\ 4a^3\ \ldots\}$
$\{e^{-an}\} = \{1\ e^{-a}\ e^{-2a}\ \ldots\}$	$\{1\ -e^{-a}\}$
$\{1 - e^{-a(n+1)}\}$	$\{1\ -1\} * \{1\ -c^{-a}\}/(1 - c^{-a})$
$\{\sin \omega(n+1)\}$	$\{1\ -2\cos\omega\ 1\}/\sin\omega$
$\{\cos \omega n\}$	$\{1\ -2\cos\omega\ 1\} * \{1 - \cos\omega\}^{-1}$
$\{e^{-a(n+1)} \sin \omega(n+1)\}$	$\{1\ -2e^{-a}\cos\omega\ c^{-2a}\}/e^{-a}\sin\omega$
$\{e^{-an} \cos \omega n\}$	$\{1\ -2e^{-a}\cos\omega\ e^{-2a}\} * \{1\ -e^{-a}\cos\omega\}^{-1}$

First we recall the special case of a matrix product where the first factor has n rows and n columns, and the second has n rows and one column only. The product is a single-column matrix of n elements. Thus

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \cdots \qquad \cdots \qquad \cdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n \end{bmatrix}$$

Now we form a column matrix $[y]$ whose elements are equal one by one to the members of the sequence $\{h\}$; we also form a column matrix $[x]$, whose early members are equal one by one to the members of the sequence $\{g\}$ as far as they go, and beyond that are zeros until there are as many elements in $[x]$ as in $[y]$. Since each member of $\{h\}$ is a linear combination of members of $\{g\}$ with a suitable set of coefficients selected from $\{f\}$, we can arrange the successive sets of coefficients, row by row, to form a square matrix such that the rules of matrix multiplication will generate the desired result. Thus, to illustrate a case where $\{f\}$ has five members, $\{g\}$ has three, and $\{h\}$ has seven, we can write:

$$\begin{bmatrix} f_0 & 0 & 0 & 0 & 0 & 0 & 0 \\ f_1 & f_0 & 0 & 0 & 0 & 0 & 0 \\ f_2 & f_1 & f_0 & 0 & 0 & 0 & 0 \\ f_3 & f_2 & f_1 & f_0 & 0 & 0 & 0 \\ f_4 & f_3 & f_2 & f_1 & f_0 & 0 & 0 \\ 0 & f_4 & f_3 & f_2 & f_1 & f_0 & 0 \\ 0 & 0 & f_4 & f_3 & f_2 & f_1 & f_0 \end{bmatrix} \begin{bmatrix} g_0 \\ g_1 \\ g_2 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} f_0g_0 \\ f_1g_0 + f_0g_1 \\ f_2g_0 + f_1g_1 + f_0g_2 \\ f_3g_0 + f_2g_1 + f_1g_2 \\ f_4g_0 + f_3g_1 + f_2g_2 \\ f_4g_1 + f_3g_2 \\ f_4g_2 \end{bmatrix} = \begin{bmatrix} h_0 \\ h_1 \\ h_2 \\ h_3 \\ h_4 \\ h_5 \\ h_6 \end{bmatrix}$$

The elements in the southeast quadrant of the f matrix could be replaced by zeros without affecting the result.

By listing parts of the sequence $\{f\}$ row by row with progressive shift as required, we have succeeded in forcing the serial product into matrix notation. This may appear to be a labored and awkward exercise, and one that sacrifices the elegant commutative property. In fact, however, matrix representation for serial products is widely used in control-system engineering and elsewhere because of the rich possibilities offered by the theory of matrices, especially infinite matrices. Moreover, in computer packages based on matrix operations, it has been found convenient to construct the convolution operation as the product of a square matrix and a column vector and to deem the time lost to computational inefficiency to be negligible.

Sequences as vectors. Just as the short sequence $\{x_1\, x_2\, x_3\}$ may be regarded as the representation of a certain vector in three-dimensional space in terms of three orthogonal components, so a sequence of n members may be regarded as representing a vector in n-dimensional space. As a rule we do not expend much effort trying to visualize the n mutually perpendicular axes along which the vector is to be resolved, but simply handle the members of the sequence according to algebraic rules which are natural extensions of those for handling vector components in three-dimensional space. Nevertheless, we borrow a whole vocabulary of terms and expressions from geometry, which lends a certain picturesqueness to linear algebra.

If the relationship between two sequences $\{x\}$ and $\{y\}$ is

$$\{a\} * \{x\} = \{y\},$$

where $\{a\}$ is some other sequence, then an equivalent statement is

$$T\,\mathbf{X} = \mathbf{Y},$$

where $\mathbf{X}$ and $\mathbf{Y}$ are the vectors whose components are $\{x\}$ and $\{y\}$, and T is an operator representing the transformation that converts $\mathbf{X}$ to $\mathbf{Y}$. The transformation consists of forming linear combinations of the components of $\mathbf{X}$ in accordance with the rules that are so concisely expressed by the serial product.

It will be recalled that the number of members of the sequence $\{y\}$ exceeds the number of members of $\{x\}$ if $\{x\}$ has a finite number of members. To avoid

the awkwardness that arises if **X** and **Y** are vectors in spaces with different numbers of dimensions, the vector interpretation is usually applied to infinite sequences $\{x\}$ and $\{y\}$. If the sequences arising in a given problem are in actual fact not infinite, it is often permissible to convert them to infinite sequences by including an infinite number of extra members all of which are zero.

As in the case of representation of sequences by column matrices, the commutative property $\{a\} * \{x\} = \{x\} * \{a\}$ is abandoned. One of the sequences entering into the serial product is interpreted as a vector, and one in an entirely different way. To reverse the roles of $\{x\}$ and $\{a\}$—for example, to talk of the vector **A**—would be unthinkable in physical fields where the vector interpretation is used. This apparent rigidification of the notation is compensated by greater generality. To illustrate this we may note first that, to the extent that matrix methods are used for dealing with vectors, there is no distinction between the representations of sequences as column matrices and as vectors. Now, examination of the square matrix in the preceding section that was built up from members of the sequence $\{f\}$ reveals that in each row the sequence $\{f\}$ was written backward, the sequence shifting one step from each row to the next. This reflects the mode of calculation of the serial product, where the sequence $\{f\}$ would be written on a movable strip of paper alongside a sequence $\{g\}$ and shifted one step after each cycle of forming products and adding. In the procedure for taking serial products there is no provision for changing the sequence $\{f\}$ from step to step. Indeed it is the essence of convolution, and of serial multiplication, that no such change occurs. But in the square-matrix formulation, it is just as convenient to express such changes as not. Thus the matrix notation allows for a general situation of which the serial product is a special case.

This more general situation is simply the general linear transformation as contrasted with the general shift-invariant linear transformation. If the indexing of a sequence is with respect to time, then serial multiplication is the most general time-invariant linear transformation, and would for example be applied in problems concerned with linear filters whose elements did not change as time went by. If, on the other hand, one were dealing with the passage of a signal through a filter containing time-varying linear elements such as motor-driven potentiometers, then the relationship between output and input could not be expressed in the form of a serial product. In such circumstances, where convolution is inapplicable, the property referred to loosely as "harmonic response to harmonic excitation" also breaks down.

CONVOLUTION BY COMPUTER

Convolution gains extra importance in Fourier analysis because of the convolution theorem which is taken up in Chapter 6, and it should be mentioned here that convolution can be based on fast transform algorithms as discussed in Chap-

ter 12. However, it is much more straightforward to code the hand calculation described in connection with Figure 3.9, and in applications to filtering long data sets this code will usually run faster than a transform method (both being written in the same language). Let `f()` and `g()` have lengths `lf` and `lg` respectively; the convolution `h()` will have length `lh` equal to `lf+lg-1`. Here is the code for direct convolution.

```
FOR I = 1 to lh
   h(I) = 0
   FOR J = MAX(1 - lg + I,1) TO MIN(lf,I)
      K = I - J + 1
      h(I) = h(I) + f(J) * g(K)
   NEXT J
NEXT I
```

THE AUTOCORRELATION FUNCTION AND PENTAGRAM NOTATION

The self-convolution of a function $f(x)$ is given by

$$f * f = \int_{-\infty}^{\infty} f(u)f(x-u)\,du.$$

Suppose, however, that prior to multiplication and integration we do not reverse one of the two component factors; then we have the integral

$$\int_{-\infty}^{\infty} f(u)f(u-x)\,du,$$

which may be denoted by $f \star f$. A single value of $f \star f$ is represented by the shaded area in Figure 3.11. A moment's thought will show that if the function f is to be displaced relative to itself by an amount x (without reversal), then the integral of the product will be the same whether x is positive or negative. In other words, if $f(x)$ is a real function, then $f \star f$ is an even function, a fact which is not

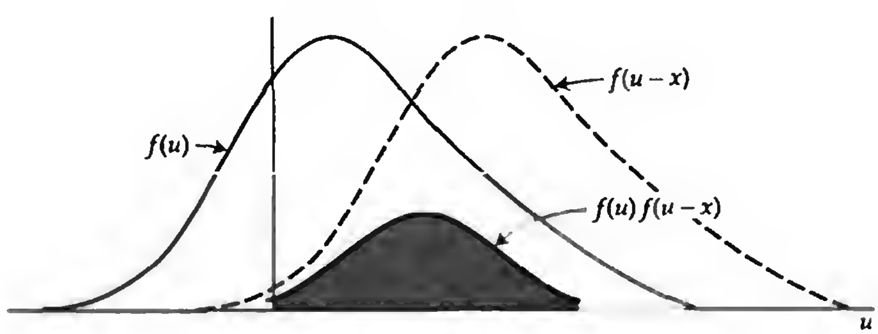


Fig. 3.11 The autocorrelation function represented by an area (shown shaded).

true in general of the convolution integral. It follows that

$$f \star f = \int_{-\infty}^{\infty} f(u)f(u-x)\,du = \int_{-\infty}^{\infty} f(u)f(u+x)\,du,$$

which, of course, is deducible from the previous expression by substitution of $w = u - x$.

It is shown in the appendix to this chapter that the value of $f \star f$ at its origin is a maximum; that is, as soon as some shift is introduced, the integral of the product falls off.

When $f(x)$ is complex it is customary to use the expression

$$\int_{-\infty}^{\infty} f(u)f^*(u-x)\,du$$

or

$$\int_{-\infty}^{\infty} f^*(u)f(u+x)\,du.$$

(Note that it is possible to place the asterisk in the wrong place and thus obtain the conjugate of the standard version.)

It is often convenient to normalize by dividing by the central value. Then we may define a quantity $\gamma(x)$ given by

$$\gamma(x) = \frac{\int_{-\infty}^{\infty} f(u)f(u+x)\,du}{\int_{-\infty}^{\infty} f(u)f(u)\,du}$$

and it is clear that

$$\gamma(0) = 1.$$

We shall refer to $\gamma(x)$ as the autocorrelation function of $f(x)$. However, it often happens that the question of normalization is unimportant in a particular application, and the character of the autocorrelation is of more interest than the magnitude; then the nonnormalized form is referred to as the autocorrelation function.

As an example, take the function $f(x)$ defined by

$$f(x) = \begin{cases} 0 & x < 0 \\ 1 - x & 0 < x < 1 \\ 0 & x > 1. \end{cases}$$

Then, for $0 < x < 1$,

$$\int_{-\infty}^{\infty} f(u)f(u+x)\,du = \int_{0}^{1-x} (1-u)[1-(u+x)]\,du$$
$$= \frac{1}{3} - \frac{x}{2} + \frac{x^3}{6}.$$

The best way to determine the limits of integration is to make a graph such as the one in Figure 3.11. Where $x > 1$, the integral is zero, and since the integral is known to be an even function of x, we have

$$\int_{-\infty}^{\infty} f(u)f(u+x)\,du = \begin{cases} \dfrac{1}{3} - \dfrac{|x|}{2} + \dfrac{|x|^3}{6} & -1 < x < 1 \\ 0 & |x| > 1. \end{cases}$$

The central value, obtained by putting $x = 0$, is $\frac{1}{3}$; hence

$$\gamma(x) = \begin{cases} 1 - \dfrac{3|x|}{2} + \dfrac{|x|^3}{2} & -1 < x < 1 \\ 0 & |x| > 1. \end{cases}$$

We note that $\gamma(0) = 1$, and as with the convolution integral, the area under the autocorrelation function may be checked (it is $\frac{1}{4}$) to verify that it is equal to the square of the area under $f(x)$.

A second example is furnished by

$$f(x) = \begin{cases} e^{-ax} & x > 0 \\ 0 & x < 0. \end{cases}$$

Then

$$\int_{-\infty}^{\infty} f(u)f(u+x)\,du = \int_{0}^{\infty} e^{-au}\,e^{-a(x+u)}\,du$$

$$= \frac{e^{-a|x|}}{2\alpha},$$

and

$$\gamma(x) = e^{-a|x|}.$$

The autocorrelation function is often used in the study of functions representing observational data, especially observations exhibiting some degree of randomness, and ingenious computing machines have been devised to carry out the integration on data in various forms. In any case, the digital computation is straightforward in principle. In the theory of such phenomena, however, as distinct from their observation and analysis, one wishes to treat functions that run on indefinitely, which often means that the infinite integral does not exist. This may always be handled by considering a segment of the function of length X and replacing the values outside the range of this segment by zero; any difficulty associated with the fact that the original function ran on indefinitely is thus removed. The autocorrelation $\gamma(x)$ of the new function $f_X(x)$, which is zero outside the finite segment, is given, according to the definition, by

$$\gamma(x) = \frac{\displaystyle\int_{-\infty}^{\infty} f_X{}^*(u)f_X(u+x)\,du}{\displaystyle\int_{-\infty}^{\infty} f_X(u)f_X{}^*(u)\,du}$$

$$= \frac{\int_{-\frac{1}{2}X}^{\frac{1}{2}X} f^*(u)f(u+x)\,du}{\int_{-\frac{1}{2}X}^{\frac{1}{2}X} f_X(u)f_X^*(u)\,du}.$$

Thus the infinite integrals are reduced to finite ones, and, of course, this is precisely what happens in fact when a calculation is made on a finite quantity of observational data.

Now if we are given a function $f(x)$ to which the definition of $\gamma(x)$ does not apply because $f(x)$ never dies out, and we have calculated $\gamma(x)$ for finite segments, we can then make the length X of the segment as long as we wish. It may happen that as X increases, the values of $\gamma(x)$ settle down to a limit; in fact, the circumstances under which this happens are of very wide interest. This limit, when it exists, will be denoted by $C(x)$, and it is given by

$$C(x) = \lim_{X\to\infty} \frac{\int_{-\frac{1}{2}X}^{\frac{1}{2}X} f(u)f(u+x)\,du}{\int_{-\frac{1}{2}X}^{\frac{1}{2}X} [f(u)]^2\,du}.$$

As autocorrelation usually arises with real functions, such as signal waveforms, allowance has not been made for complex $f(\ )$.

Exercise. Show that when this expression is evaluated for a real function for which $\gamma(x)$ exists, then $C(x) = \gamma(x)$.

Since the limiting autocorrelation C is identical with γ in cases where γ is applicable, it would be logical to take C as defining the autocorrelation function, and this is often done. In the discussion that follows, however, we shall understand the term "autocorrelation function" to include the operation of passing to a limit only where that is necessary.

As an example of a function that does not die out and for which the infinite integral does not exist, consider a time-varying signal which is a combination of three sinusoids of arbitrary amplitudes and phases, given by

$$V(t) = A\sin(\alpha t + \phi) + B\sin(\beta t + \chi) + C\sin(\gamma t + \psi).$$

Then

$$\begin{aligned}\int_{-\frac{1}{2}T}^{\frac{1}{2}T} V(t')V(t'+t)\,dt' = \int_{-\frac{1}{2}T}^{\frac{1}{2}T} [&A\sin(\alpha t' + \phi)\,A\sin(\alpha t' + \alpha t + \phi)\\ &+ B\sin(\beta t' + \chi)\,B\sin(\beta t' + \beta t + \chi)\\ &+ C\sin(\gamma t' + \psi)\,C\sin(\gamma t' + \gamma t + \psi)\\ &+ \text{cross-product terms}]\,dt'\end{aligned}$$

$$= \int_{-\frac{1}{2}T}^{\frac{1}{2}T} \{A^2[\cos \alpha t - \cos(2\alpha t' + \alpha t + 2\phi)] + B^2[\cos \beta t - \cos(2\beta t' + \beta t + 2\chi)] + C^2[\cos \gamma t - \cos(2\gamma t' + \gamma t + 2\psi)] + \text{cross-product terms}\}\, dt'$$

$$= A^2T \cos \alpha t + B^2T \cos \beta t + C^2T \cos \gamma t + F(t,T) + G(t,T),$$

where F stands for oscillatory terms and G for terms arising from cross products. As T increases, F and G become negligible with respect to the three leading terms, which increase in proportion to T. Hence

$$\frac{\int_{-\frac{1}{2}T}^{\frac{1}{2}T} V(t')V(t' + t)\, dt'}{\int_{-\frac{1}{2}T}^{\frac{1}{2}T} [V(t')]^2\, dt'} = \frac{A^2T \cos \alpha t + B^2T \cos \beta t + C^2T \cos \gamma t + F(t,T) + G(t,T)}{A^2T + B^2T + C^2T + F(0,T) + G(0,T)}$$

and

$$C(t) = \lim_{T\to\infty} \frac{\int_{-\frac{1}{2}T}^{\frac{1}{2}T} V(t')V(t' + t)\, dt'}{\int_{-\frac{1}{2}T}^{\frac{1}{2}T} [V(t')]^2\, dt'}$$

$$= \frac{1}{A^2 + B^2 + C^2} (A^2 \cos \alpha t + B^2 \cos \beta t + C^2 \cos \gamma t).$$

Note that the limiting autocorrelation function $C(t)$ is a superposition of three *cosine* functions at the same frequencies as contained in the signal $V(t)$, but with different relative amplitudes, and that $C(0) = 1$.

Since the principal terms in the numerator of $\gamma(x)$ increase in proportion to X, another way to obtain a nondivergent result is to divide by X before proceeding to the limit. The expression

$$\lim_{X\to\infty} \frac{1}{X} \int_{-\frac{1}{2}X}^{\frac{1}{2}X} f(u)f(u + x)\, du$$

is not equal to unity at its origin but is generally referred to simply as the "autocorrelation function" because it becomes the same as $C(x)$ if divided by its central value (provided the central value is not zero). It may therefore be regarded as a nonnormalized form of $C(x)$.

When time is the independent variable it is customary to refer to the time average of the product $V(t)V(t + \tau)$, and write

$$\langle V(t)V(t + \tau)\rangle \equiv \lim_{T\to\infty} \frac{1}{T} \int_{-\frac{1}{2}T}^{\frac{1}{2}T} V(t)V(t + \tau)\, dt,$$

where the operation of time averaging is denoted by angular brackets according to the definition

$$\langle \ldots \rangle \equiv \lim_{T\to\infty} \frac{1}{T} \int_{-\frac{1}{2}T}^{\frac{1}{2}T} \ldots \, dt.$$

Using this notation for the example worked, we have

$$\langle V(t)V(t+\tau)\rangle = A^2 \cos \alpha\tau + B^2 \cos \beta\tau + C^2 \cos \gamma\tau,$$

which, of course, is a little simpler than the normalized expression. It should be noticed, however, that we do not get a useful result in cases for which γ exists.

A conspicuous feature of C in this example is the absence of any trace of the original phases. Hence, autocorrelation is not a reversible process; it is not possible to get back from the autocorrelation function to the original function from which it was derived. Autocorrelation thus involves a loss of information. In some branches of physics, such as radio interferometry and X-ray diffraction analysis of substances, it is easier to observe the autocorrelation of a desired function than to observe the function itself, and a lot of ingenuity is expended to fill in the lost information.

The character of the lost information can be seen by considering a cosinusoidal function of x. When this function is displaced relative to itself, multiplied with the unshifted function, and the result integrated, clearly the result will be the same as for a sinusoidal function of the same period and amplitude. Furthermore, the result will be the same for any harmonic function of x, with the same period and amplitude, and arbitrary origin of x. Thus the autocorrelation function does not reveal the phase of a harmonic function. Now, if a function is composed of several harmonic waves present simultaneously, then when it is displaced, multiplied, and integrated, the result can be calculated simply by considering the different periods one at a time. This is possible because the product of harmonic variations of different frequencies integrates to a negligible quantity. Consequently, each of the periodic waves may be slid along the x axis into any arbitrary phase, without affecting the autocorrelation. In particular, all the components may be shifted until they become cosine components, thus generating an even function which, among all possible functions with the same autocorrelation, possesses a certain uniqueness.

As autocorrelation usually arises with real functions, such as signal waveforms, allowance has not been made for complex f. The autocorrelation of a complex function $f(x)$ can be treated as a special case of cross correlation.

THE TRIPLE CORRELATION

Ordinary autocorrelation of a waveform abandons information about the direction of time, which can be important. For example, random noise that is statistically the same if the record is reversed can be obtained from a physical source; the same noise, passed through a simple RC-filter with impulse response $\exp(-t/RC)H(t)$ can be thought of as a superposition of exponential tails all

imprinted with the direction of time flow. Therefore the filtered noise may be statistically different if played backwards. The autocorrelation function does not allow us to investigate this because it tells nothing about the arrow of time. But there is a higher order property that does and has become important (Weigelt, 1991), namely the triple correlation. It is a function $U(x_1, x_2)$ of two variables defined as

$$U(x_1, x_2) = \int_{-\infty}^{\infty} f(x)f(x + x_1)f(x + x_2)\,dx.$$

For example, the triple correlations of {1 2 3} and {3 2 1} respectively are

$$\begin{bmatrix} 0 & 0 & 3 & 6 & 9 \\ 0 & 6 & 14 & 22 & 6 \\ 9 & 22 & \mathbf{36} & 14 & 3 \\ 6 & 14 & 22 & 6 & 0 \\ 3 & 6 & 9 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 & 0 & 9 & 6 & 3 \\ 0 & 6 & 22 & 14 & 6 \\ 3 & 14 & \mathbf{36} & 22 & 9 \\ 6 & 22 & 14 & 6 & 0 \\ 9 & 6 & 3 & 0 & 0 \end{bmatrix},$$

and have come out differently, one being the transpose of the other.

THE CROSS CORRELATION

The cross correlation $f(x)$ of two real functions $g(x)$ and $h(x)$ is defined as

$$f(x) = g \star h = \int_{-\infty}^{\infty} g(u - x)h(u)\,du.$$

It is thus similar to the convolution integral except that the component $g(u)$ is simply displaced to $g(u - x)$, without reversal. To denote this operation of g on h we use a five-pointed star, or pentagram, instead of the asterisk that denotes convolution. While cross correlation is slightly simpler than convolution, its properties are less simple. Thus, whereas $g * h = h * g$, the cross-correlation operation of h on g is not the same as that of g on h. It is a good idea to read $g \star h$ as 'g scans h.' As defined above g scans h to produce f, since h remains stationary while g moves with changing x. We now show that

$$g \star h \neq h \star g.$$

By change of variables it can be seen that

$$f(x) = g \star h = \int_{-\infty}^{\infty} g(u - x)h(u)\,du = \int_{-\infty}^{\infty} g(u)h(u + x)\,du$$

$$f(-x) = h \star g = \int_{-\infty}^{\infty} h(u - x)g(u)\,du = \int_{-\infty}^{\infty} h(u)g(u + x)\,du;$$

$$g * h = g(-) \star h.$$

As in the case of the autocorrelation function, the cross correlation is often normalized so as to be equal to unity at the origin, and when appropriate the average $\langle g(u - x)h(u)\rangle$ is used instead of the infinite integral. When the functions are complex it is customary to define the (complex) cross-correlation function as g^* scans h, where g^* is the complex conjugate of g:

$$g^* \star h = \int_{-\infty}^{\infty} g^*(u - x)h(u)\, du = \int_{-\infty}^{\infty} g^*(u)h(u + x)\, du.$$

As a special case the complex autocorrelation of complex f is $f^* \star f$, which requires care because it is possible to put the asterisk in the wrong place and thus obtain the conjugate of the standard version.

THE ENERGY SPECTRUM

We shall refer to the squared modulus of a transform as the energy spectrum; that is, $|F(s)|^2$ is the energy spectrum of $f(x)$. The term is taken directly from the physical fields where it is used. It will be seen that there is not a one-to-one relationship between $f(x)$ and its energy spectrum, for although $f(x)$ determines $F(s)$ and hence also $|F(s)|^2$, it would be necessary to have[1] pha $F(s)$ as well as $|F(s)|$ in order to reconstitute $f(x)$. Knowledge of the energy spectrum thus conveys a certain kind of information about $f(x)$ which says nothing about the phase of its Fourier components. It is the kind of information about an acoustical waveform which results from measuring the sound intensity as a function of frequency.

The information lost when only the energy spectrum can be given is of precisely the same character as that which is lost when the autocorrelation function has to do duty for the original function. The autocorrelation theorem, to be given later, expresses this equivalence.

When $f(x)$ is real, as it would be if it represented a physical waveform, the energy spectrum is an even function and is therefore fully determined by its values for $s \geqslant 0$. To stress this fact, the term "positive-frequency energy spectrum" may be used to mean the part of $|F(s)|^2$ for which $s \geqslant 0$.

Since $|F(s)|^2$ has the character of energy density measured per unit of s, it would have to become infinite if a nonzero amount of energy were associated with a discrete value of s. This is the situation with an infinitely narrow spectral

[1]The phase angle of the complex quantity $F(s)$ is written pha $F(s)$ and is defined as follows: if there is a complex variable $z = r \exp i\theta$, then pha $z = \theta$.

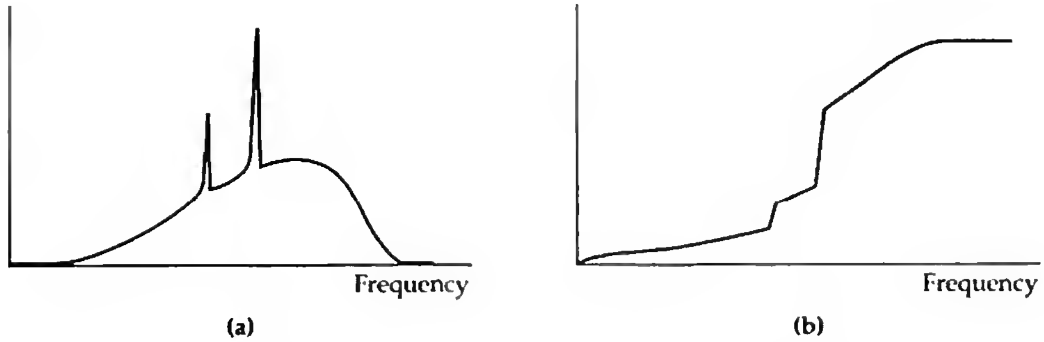


Fig. 3.12 **Spectra of X radiation from molybdenum: (a) power spectrum; (b) cumulative spectrum.**

line. Now we can consider a cumulative distribution function which gives the amount of energy in the range 0 to s:

$$\int_0^s |F(s)|^2 \, ds.$$

Any spectral *lines* would then appear as finite discontinuities in the cumulative energy spectrum as suggested by Fig. 3.12, and some mathematical convenience would be gained by using the cumulative spectrum in conjunction with Stieltjes integral notation. The convenience is especially marked when it is a question of using the theory of distributions for some question of rigor. However, the matter is purely one of notation, and in cases where we have to represent concentrations of energy within bands much narrower than can be resolved in the given context, we shall use the delta-symbol notation described later.

APPENDIX

We prove that the autocorrelation of the real nonzero function $f(x)$ is a maximum at the origin, that is,

$$\int_{-\infty}^{\infty} f(u)f(u+x)\,du \leq \int_{-\infty}^{\infty} [f(u)]^2\,du.$$

Let ϵ be a real number. Then, if $x \neq 0$,

$$\int_{-\infty}^{\infty} [f(u) + \epsilon f(u+x)]^2\,du > 0$$

and $$\int_{-\infty}^{\infty} [f(u)]^2\,du + 2\epsilon\int_{-\infty}^{\infty} f(u)f(u+x)\,du + \epsilon^2\int_{-\infty}^{\infty} [f(u+x)]^2\,du > 0;$$

that is,
$$a\epsilon^2 + b\epsilon + c > 0,$$
where
$$a = c = \int_{-\infty}^{\infty} [f(u)]^2\, du$$
$$b = 2\int_{-\infty}^{\infty} f(u)f(u + x)\, du.$$

Now, if the quadratic expression in ϵ may not be zero, that is, if it has no real root, then
$$b^2 - 4ac \leqslant 0.$$
Hence in this case $b/2 \leqslant a$, or
$$\frac{\int_{-\infty}^{\infty} f(u)f(u + x)\, du}{\int_{-\infty}^{\infty} [f(u)]^2\, du} \leqslant 1.$$
The equality is achieved at $x = 0$; consequently the autocorrelation function can nowhere exceed its value at the origin. The argument is the one used to establish the Schwarz inequality and readily generalizes to give the similar result for the complex autocorrelation function.

Exercise. Extend the argument with a view to showing that the equality cannot be achieved by any value of x save zero.

■

BIBLIOGRAPHY

Bracewell, R. N.: "Two-Dimensional Imaging," Prentice-Hall, Englewood Cliffs, N.J., 1995.

Bracewell, R. N., and J. A. Roberts: "Aerial Smoothing in Radio Astronomy," *Australian J. Physics*, vol. 7, pp. 615–640, 1954.

Pettit, R. H.: "Pulse Sequences with Good Autocorrelation Properties," *Microwave Journal*, vol. 10, pp. 63–67, 1967.

Weigelt, G.: *Progress in Optics*, vol. 29, p. 293, 1991.

■

PROBLEMS

1. Calculate the following serial products, checking the results by summation. Draw graphs to illustrate.
   (*a*) {6 9 17 20 10 1} * {3 8 11}
   (*b*) {1 1 1 1 1} * {1 1 1 1}
   (*c*) {1 4 2 3 5 3 3 4 5 7 6 9} * {1 1}
   (*d*) {1 4 2 3 5 3 3 4 5 7 6 9} * {$\frac{1}{2}$ $\frac{1}{2}$}

(*e*) $\{1\,4\,2\,3\,5\,3\,3\,4\,5\,7\,6\,9\} * \{1\,2\,1\}$
(*f*) $\{1\,4\,2\,3\,5\,3\,3\,4\,5\,7\,6\,9\} * \{\frac{1}{4}\,\frac{1}{2}\,\frac{1}{4}\}$
(*g*) $\{1\,2\,3\,7\,12\,19\,21\,22\,18\,13\,7\,5\,3\,2\,1\} * \{1\ {-1}\}$
(*h*) $\{1\,2\,3\,7\,12\,19\,21\,22\,21\,18\,13\,7\,5\,3\,2\,1\} * \{1\,1\,1\,1\,\ldots\}$
(*i*) $\{1\,1\} * \{1\,1\} * \{1\,1\}$
(*j*) $\{1\,1\,0\,0\,1\,0\,1\} * \{1\,1\,0\,0\,1\,0\,1\}$
(*k*) $\{1\,1\,0\,0\,1\,0\,1\} * \{1\,0\,1\,0\,0\,1\,1\}$
(*l*) $\{a\,b\,c\,d\,e\} * \{e\,d\,c\,b\,a\}$
(*m*) $\{1\,3\,1\} * \{1\,0\,0\,0\,0\,0\,\pm 1\}$
(*n*) $\{1\,3\,1\} * \{1\,2\,2\}$
(*o*) Multiply 131 by 122
(*p*) Multiply 10,301 by 10,202
(*q*) $\{1\,8\,1\} * \{1\,2\,2\}$
(*r*) Comment on the smoothness of your results in *d* and *f* relative to the longer of the two given sequences.
(*s*) Consider the result of *i* in conjunction with Pascal's pyramid of binomial coefficients.
(*t*) Seek longer sequences with the same property you discovered in *k*.
(*u*) Contemplate *j*, *k*, *l*, *m*, and *n* with a view to discerning what leads to serial products which are even.
(*v*) Master the implication of *n*, *o*, *p*, and *q*, and design a mechanical desk computer to perform serial multiplication.

2. Derive the following results, where $H(x)$ is the Heaviside unit step function (Chapter 4):

$$\begin{aligned}
x^2H(x) * e^xH(x) &= (2e^x - x^2 - 2x - 2)H(x)\\
[\sin x\, H(x)]^{*2} &= \tfrac{1}{2}(\sin x - x\cos x)H(x)\\
[(1-x)H(x)] * [e^xH(x)] &= xH(x)\\
H(x) * [e^xH(x)] &= (e^x - 1)H(x)\\
[e^xH(x)]^{*2} &= xe^xH(x)\\
[e^xH(x)]^{*3} &= \tfrac{1}{2}x^2e^xH(x)
\end{aligned}$$

3. Prove the commutative property of convolution, that is, that $f * g \equiv g * f$.

4. Prove the associative rule $f * (g * h) \equiv (f * g) * h$.

5. Prove the distributive rule for addition $f * (g + h) \equiv f * g + f * h$.

6. The function f is the convolution of g and h. Show that the self-convolutions of f, g, and h are related in the same way as the original functions. ▷

7. If $f = g \star h$, show that $f \star f = (g \star g) \star (h \star h)$.

8. Show that if a is a constant, $a(f * g) = (af) * g = f * (ag)$.

9. Establish a theorem involving $f(g * h)$. ▷

10. Prove that the autocorrelation function is hermitian, that is, that $C(-u) = C^*(u)$, and hence that when the autocorrelation function is real it is even. Note that if the autocor-

relation function is imaginary it is also odd; give some thought to devising a function with an odd autocorrelation function.

11. Prove that the sum and product of two autocorrelation functions are each hermitian. ▷

12. Alter the origin of $f(x)$ until $f * f|_0$ is a maximum. Investigate the assertion that the new origin defines an axis of maximum symmetry, making any necessary modification. Investigate the merits of the parameter

$$\frac{f*f|_0}{f \star f|_0}$$

to be considered a measure of "degree of evenness."

13. Show that if $f(x)$ is real,

$$\int_{-\infty}^{\infty} f(x)f(-x)\,dx = \int_{-\infty}^{\infty} [E(x)]^2\,dx - \int_{-\infty}^{\infty} [O(x)]^2\,dx,$$

and note that the left-hand side is the central value of the self-convolution of $f(x)$; that is, $f * f|_0$.

14. Find reciprocal sequences for $\{1\ 3\ 3\ 1\}$ and $\{1\ 4\ 6\ 4\ 1\}$.

15. Find reciprocal sequences for $\{1\ 1\}$ and $\{1\ 1\ 1\}$.

16. Establish a general procedure for finding the reciprocal of finite or semi-infinite sequences and test it on the following cases:

$$\{64\ 32\ 16\ 8\ 4\ 2\ 1\ \ldots\}$$

$$\{64\ 64\ 48\ 32\ 20\ 12\ 7\ 4\ \ldots\}$$

$$\left\{1\quad e^{-1/10}\cos\left(\frac{\pi}{10}\right)\quad e^{-2/10}\cos\left(\frac{2\pi}{10}\right)\quad \ldots\ e^{-n/10}\cos\left(\frac{n\pi}{10}\right)\quad \ldots\right\}$$

17. Find approximate numerical values for a function $f(x)$ such that

$$f(x) * [e^{-x}H(x)]$$

is zero when evaluated numerically by serial multiplication of values taken at intervals of 0.2 in x, except at the origin. Normalize $f(x)$ so that its integral is approximately unity. ▷

18. The cross correlation $g \star h$ is to be normalized to unity at its maximum value. It is argued that

$$0 \leqslant \int [g(u) - h(u+x)]^2\,du = \int g^2\,du - 2\int g(u)h(u+x)\,du + \int h^2\,du,$$

and therefore that

$$\int g(u)h(u+x)\,du \leqslant \tfrac{1}{2}\int g^2\,du + \tfrac{1}{2}\int h^2\,du = M.$$

Consequently $(g \star h)/M$ is the desired quantity. Correct the fallacy in this argument.

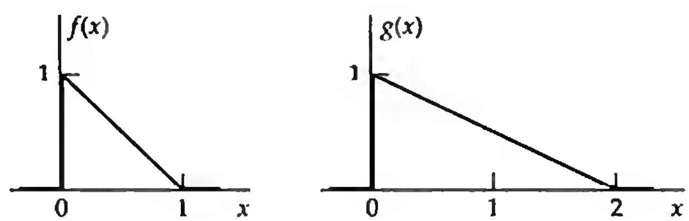
19. Barker code. Calculate the autocorrelation sequence of

$$\{1\ 1\ 1\ 1\ 1\ -1\ -1\ 1\ 1\ -1\ 1\ -1\ 1\}.$$

This sequence is known as the Barker code of length 13 (Pettit, 1967). By coin tossing, establish a similar 13-element sequence at random and calculate its autocorrelation.

20. Convolution done analytically.

(*a*) Make a graph of the convolution $h(x)$ of the given functions $f(x)$ and $g(x)$, labeling both axes with numerical values.



(*b*) Label any interesting points of $h(x)$ with letters $A, B, C, \ldots$ and make a table of values such as the following:

Interesting point	x	$h(x)$
A		
B		
C		
...		

21. Self-convolution of sinc function. Let $f(x) = \operatorname{sinc}(x+2) + \operatorname{sinc}(x-2)$. What is the self-convolution of $f(x)$?

22. Convolution. In the integral

$$\int_{-\infty}^{\infty} f(x-u)g(u)\,du,$$

make the substitution

$$u = x - a,$$

where a is a constant. Then

$$\int_{-\infty}^{\infty} f(x-u)g(u)\,du = \int_{-\infty}^{\infty} f(a)g(x-a)\,dx = f(a)\int_{-\infty}^{\infty} g(x)\,dx.$$

Is this correct? If not, where is the fallacy in the derivation?

23. Optical sound track. The optical sound track on old motion-picture film has a breadth b, and it is scanned by a slit of width w. With appropriate normalization, we may say that the scanning introduces convolution by a rectangle function of unit height and width w. In a certain movie theater the projectionist clumsily dropped the whole projector on the floor and after that the slit was always inclined at a small angle ϵ to the striations on the sound track instead of making an angle of zero with them.
(*a*) What function now describes the convolution that takes place?
(*b*) Describe qualitatively the effect on the sound reproduction.

24. Numerical convolution compared with analytic. The exponential function e^{-x} may be represented discretely by the sequence

$$\{a\} = \{1 \quad 0.368 \quad 0.135 \quad 0.050 \quad \ldots\}$$

(*a*) Calculate the serial product or "autocorrelation sum"

$$\sum_{j=0,1,\ldots} a_j a_{i+j}$$

and make an accurate graph.
(*b*) Calculate the "autocorrelation function"

$$R(\tau) = \int_{-\infty}^{\infty} f(x)f(x+\tau)\,dx$$

when

$$f(x) = \begin{cases} e^{-x} & x > 0 \\ 0 & x < 0, \end{cases}$$

and superimpose a graph of $R(\tau)$ on the previous graph.
(*c*) Naturally the continuous graph of $R(\tau)$ does not pass exactly through the points calculated from part *a*. Discuss the discrepancy in terms of round-off error, normalization, or any other effects which you think may account for the disagreement.

25. Two-dimensional convolution. The autocorrelation of two disks of unit diameter arises in the theory of optical instruments with circular apertures (and describes the two-dimensional transfer function). It is known as the Chinese hat function. Show that

$$\Pi(r) ** \Pi(r) = \text{chat}\, r = \tfrac{1}{2}[\cos^{-1} r - r(1-r^2)^{\frac{1}{2}}]\Pi\left(\frac{r}{2}\right). \triangleright$$

26. Two-dimensional autocorrelation. The two-dimensional autocorrelation function of $\Pi(x)\Pi(y)$ is $\Lambda(x)\Lambda(y)$. The central value is unity.
(*a*) Verify that the contour $\Lambda(x)\Lambda(y) = \epsilon$, where ϵ is small compared with unity, is approximately square.
(*b*) Verify that the contour $\Lambda(x)\Lambda(y) = 1 - \delta$, where δ is small compared with unity, is also approximately square.
(*c*) Choosing $\delta = \epsilon$, would you say that the two contours are equally square?

27. Autocorrelation of a convolution. Show that

$$(f * g) \star (f * g) = (f \star f) * (g \star g).$$

28. Deconvolution. Three sequences are related by convolution as follows:

$$\{a_0\, a_1\, a_2\, \ldots\} * \{b_0\, b_1\, b_2\, \ldots\} = \{c_0\, c_1\, c_2\, \ldots\}.$$

If the sequences $\{b_k\}$ and $\{c_k\}$ are given, show that the rule for inverting the convolution to obtain $\{a_k\}$ is

$$a_k = b_0^{-1}\left(c_k - \sum_{j=0}^{k-1} a_j b_{k-j}\right).$$

Write a computer subprogram for inverting convolution.

29. Transmission line echoes. When a current impulse is injected into a transmission line that is short-circuited at the far end the voltage appearing at the input terminals is a sequence of equispaced impulses with coefficients $\{v\} = \{1\,2\,2\,2\,\ldots\}$. The current response if an impulse of voltage is applied is $\{i\} = \{1\ \ -2\ \ 2\ -2\ \ 2\ \ldots\}$. (a) Show that $\{v\} * \{i\} = \{1\}$. (b) Find two functions $v(t)$ and $i(t)$ such that $v(t)\mathrm{III}(t)$ and $i(t)\mathrm{III}(t)$ respectively evoke the impulse trains. ▷

30. Deconvolution. Let $\{b_k\}$ be an unknown sequence and let $\{a_k\} = \{1\,1\,1\,1\,1\,1\}$. Let $\{p_k\}$ be the periodic sequence of which one period is $\{0\,1\,1\,-1\,-1\,0\}$. Verify that $\{a_k\} * \{p_k\} = \{0\}$. As in the previous problem, let $\{c_k\} = \{a_k\} * \{b_k\}$ be given. Is it not true that convolving $\{a_k\}$ with $\{b_k\} + \{p_k\}$ will also yield $\{c_k\}$? If this is so, how can the deconvolution rule given above recover the unknown $\{b_k\}$? What, in fact, will the deconvolution algorithm produce? ▷

31. Speed of convolution. Specify a long data sequence f with N elements, smooth it by convolution with $g = \{1\,4\,6\,4\,1\}$, and obtain the elapsed time. We know that $f * g = g * f$, but does elapsed time depend on which sequence comes first? Does the time become proportional to N^2 as N increases? ▷

4

Notation for Some Useful Functions

Many useful functions in Fourier analysis have to be defined piecewise because of abrupt changes. For example, we may consider the function $f(x)$ such that

$$f(x) = \begin{cases} 0 & x < 0 \\ x & 0 \leq x \leq 1 \\ 1 & x > 1. \end{cases}$$

This function, though simple in itself, is awkwardly expressed in comparison with a function such as, for example, $1 + x^2$, whose algebraic expression compactly states, over the infinite range of x, the arithmetical operations by which it is formed. For many mathematical purposes a function which is piecewise analytic is not simple to deal with, but for physical purposes, a "sloping step function," or "ramp-step function," may be at least as simple as a smoother function.

Fourier himself was concerned with the representation of functions given graphically, and according to English mathematician and historian E. W. Hobson "was the first fully to grasp the idea that a single function may consist of detached portions given arbitrarily by a graph."

To regain compactness and clarity of notation, we introduce a number of simple functions embodying various kinds of abrupt behavior. Also included here is a section dealing with sinc x, the important interpolating function, which is the transform of a discontinuous function, and some reference material on notations for the Gaussian function.

RECTANGLE FUNCTION OF UNIT HEIGHT AND BASE, $\Pi(x)$

The rectangle function of unit height and base, which is illustrated in Fig. 4.1, is defined by

$$\Pi(x) = \begin{cases} 0 & |x| > \frac{1}{2} \\ (\frac{1}{2} & |x| = \frac{1}{2}) \\ 1 & |x| < \frac{1}{2}. \end{cases}$$

It provides simple notation for segments of functions which have simple expressions, for example, $f(x) = \Pi(x) \cos \pi x$ is compact notation for

$$f(x) = \begin{cases} 0 & x < -\frac{1}{2} \\ \cos \pi x & -\frac{1}{2} < x < \frac{1}{2} \\ 0 & \frac{1}{2} < x \end{cases}$$

(see Fig. 4.2). We may note that $h\Pi[(x - c)/b]$ is a displaced rectangle function of height h and base b, centered on $x = c$ (see Fig. 4.3). Hence, purely by multiplication by a suitably displaced rectangle function, we can select, or gate, any segment of a given function, with any amplitude, and reduce the rest to zero.

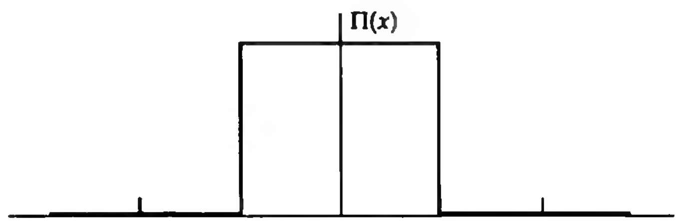


Fig. 4.1 The rectangle function of unit height and base, $\Pi(x)$.

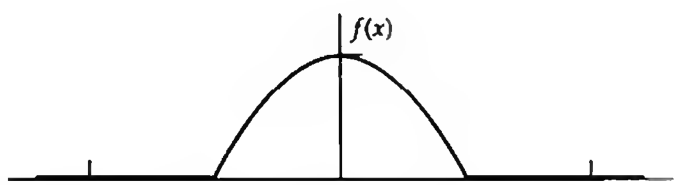


Fig. 4.2 A segmented function expressed by $\Pi(x) \cos \pi x$.

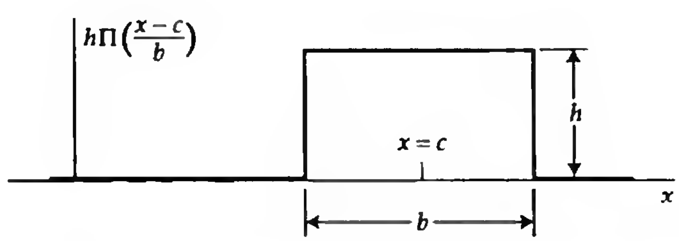


Fig. 4.3 A displaced rectangle function of arbitrary height and base expressed in terms of $\Pi(x)$.

The rectangle function also enters through convolution into expressions for running means, and, of course, in the frequency domain multiplication by the rectangle function is an expression of ideal low-pass filtering. It is important in the theory of convergence of Fourier series, where it is generally known as Dirichlet's discontinuous factor. The notation rect x is in common use as a slightly less compact alternative to $\Pi(x)$ and is useful in conversation. Terms such as gate function, window function, and boxcar function are also in use.

For reasons explained below, it is almost never important to specify the values at $x = \pm\frac{1}{2}$, that is, at the points of discontinuity, and we shall normally omit mention of those values. Likewise, it is not necessary or desirable to emphasize the values $\Pi(\pm\frac{1}{2}) = \frac{1}{2}$ in graphs; it is preferable to show graphs which are reminiscent of high-quality oscillograms (which, of course, would never show extra brightening halfway up the discontinuity).

TRIANGLE FUNCTION OF UNIT HEIGHT AND AREA, $\Lambda(x)$

By definition,

$$\Lambda(x) = \begin{cases} 0 & |x| > 1 \\ 1 - |x| & |x| < 1. \end{cases}$$

This function, which is illustrated in Fig. 4.4, gains its importance largely from being the self-convolution of $\Pi(x)$, but it has other uses—for example, in giving compact notations for polygonal functions (continuous functions consisting of linear segments).

Note that $h\Lambda(x/\frac{1}{2}b)$ is a triangle function of height h, base b, and area $\frac{1}{2}hb$.

VARIOUS EXPONENTIALS AND GAUSSIAN AND RAYLEIGH CURVES

Figure 4.5 shows, from left to right, a rising exponential, a falling exponential, a truncated falling exponential, and a double-sided falling exponential.

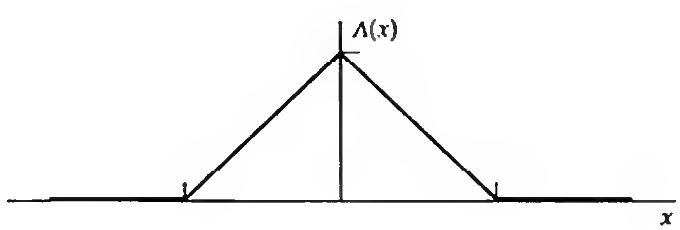


Fig. 4.4 The triangle function of unit height and area, $\Lambda(x)$.

Figure 4.6 shows the Gaussian function exp $(-\pi x^2)$. Of the various ways of normalizing this function, we have chosen one in which both the central ordinate and the area under the curve are unity, and certain advantages follow from this choice. The Fourier transform of the Gaussian function is also Gaussian, and proves to be normalized in precisely the same way under our choice. In statistics the Gaussian distribution is referred to as the "normal (error) distribution with zero mean" and is normalized so that the area and the standard deviation are unity. We may use the term "probability ordinate" to distinguish the form

$$\frac{1}{(2\pi)^{\frac{1}{2}}} e^{-\frac{1}{2}x^2}.$$

When the standard deviation is σ, the probability ordinate is

$$\frac{1}{\sigma(2\pi)^{\frac{1}{2}}} e^{-x^2/2\sigma^2},$$

and the area under the curve remains unity. The central ordinate is equal to $0.3989/\sigma$. Prior to the strict standardization now prevailing in statistics, the error integral erf x was introduced as

$$\operatorname{erf} x = \frac{2}{\pi^{\frac{1}{2}}} \int_0^x e^{-t^2}\, dt.$$

The complementary error integral erfc x is defined by

$$\operatorname{erfc} x = 1 - \operatorname{erf} x.$$

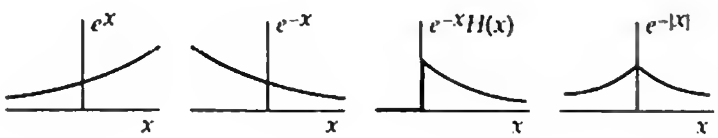


Fig. 4.5 **Various exponential functions.**

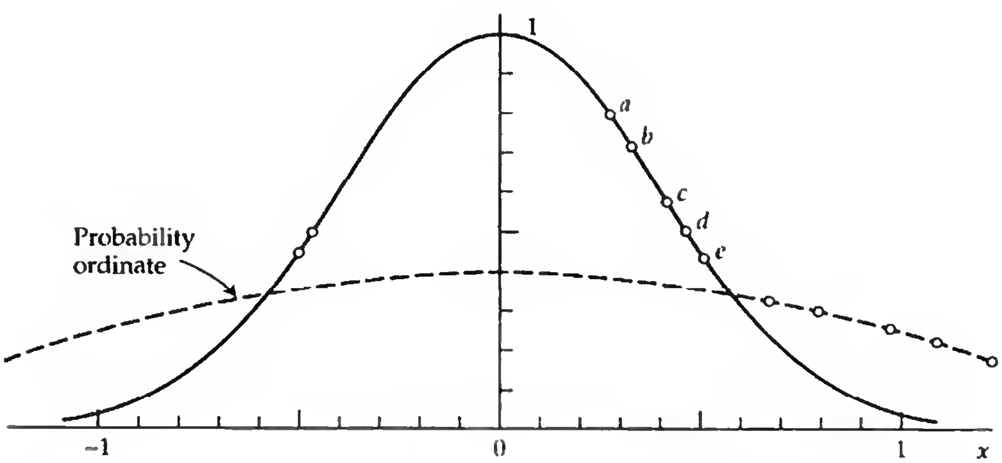


Fig. 4.6 **The Gaussian function exp $(-\pi x^2)$ and the probability ordinate $(2\pi)^{-\frac{1}{2}} \exp(-\frac{1}{2}x^2)$.**

The probability integral $\alpha(x)$, now widely tabulated, is

$$\alpha(x) = \frac{1}{(2\pi)^{\frac{1}{2}}} \int_{-x}^{x} e^{-\frac{1}{2}t^2}\, dt = \operatorname{erf} \frac{x}{2^{\frac{1}{2}}}.$$

In this volume we use $\exp(-\pi x^2)$ extensively because of its symmetry under the Fourier transformation. Its integral is related to erf x and $\alpha(x)$ as follows:

$$\int_0^x e^{-\pi t^2}\, dt = \tfrac{1}{2} \operatorname{erf} \pi^{\frac{1}{2}} x = \tfrac{1}{2}\alpha[(2\pi)^{\frac{1}{2}}x]$$

$$\int_{-\infty}^x e^{-\pi t^2}\, dt = \tfrac{1}{2} + \tfrac{1}{2} \operatorname{erf} \pi^{\frac{1}{2}} x = \tfrac{1}{2} + \tfrac{1}{2}\alpha[(2\pi)^{\frac{1}{2}}x].$$

The customary dispersion parameters of $\exp(-\pi x^2)$ are (see Fig. 4.6) as follows:

(*a*) Probable error	$= 0.2691$	$= 0.6745\sigma$
(*b*) Mean absolute error (mean of $\lvert x \rvert$) $= \pi^{-1}$	$= 0.3183$	$= 0.7979\sigma$
(*c*) Standard deviation (mean of x^2) $= (2\pi)^{-\frac{1}{2}}$	$= 0.3989$	$= \sigma$
(*d*) Width to half-peak	$= 0.9394$	$= 2.355\sigma$
(*e*) Equivalent width	$= 1.0000$	$= 2.5066\sigma$

The probable error, or semi-interquartile range, characterizes the middle 50 percent of a set of measurements. A statistical report such as "the students have an average height of 165±10 cm" means that half the students are in the range 155 to 175 cm, unless some other measure of spread is specified. For comparison, those departing from the mean of a normal distribution by less than one standard deviation are expected to number 62.27 percent, by less than the mean absolute error 57.51 percent, by less than the semiwidth to half peak 76.10 percent, and by less than the semi-equivalent width 78.99 percent.

In two dimensions the Gaussian distribution generalizes to

$$e^{-\pi(x^2+y^2)},$$

again with symmetry under the Fourier transformation, with unit central ordinate, and with unit volume. The version used in statistics, for arbitrary standard deviations σ_x and σ_y, is

$$\frac{1}{2\pi\sigma_x\sigma_y} e^{-x^2/2\sigma_x^2 - y^2/2\sigma_y^2}.$$

Under conditions of circular symmetry, and putting $x^2 + y^2 = r^2$, the two-dimensional probability ordinate becomes

$$\frac{1}{2\pi\sigma^2} e^{-r^2/2\sigma^2}.$$

The probability $R(r)\, dr$ of finding the radial distance in the range r to $r + dr$ is $2\pi r\, dr$ times the above expression; hence

$$R(r) = \frac{r}{\sigma^2} e^{-r^2/2\sigma^2}.$$

This is referred to as Rayleigh's distribution, since it occurred in the famous problem of the drunkard's walk discussed by Rayleigh. In a simple version of the problem the drunkard always falls down after taking one step, and the direction of each step bears no relation to the previous step. After a long time has elapsed, the probability of finding him at (x,y) is a two-dimensional Gaussian function (according to which he is more likely to be at the origin than elsewhere), and the probability of finding him at a distance r from the origin is given by a Rayleigh distribution. Since the Rayleigh distribution has a peak that is not at the origin, the above statements may appear contradictory. If they do, the reader will find it instructive to contemplate the matter further.

Here are some integrals and derivatives often needed for checking:

$$\int_{-\infty}^{\infty} e^{-\pi x^2}\,dx = 1 \qquad \int_{-\infty}^{\infty} e^{-x^2}\,dx = \pi^{\frac{1}{2}} \qquad \int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} = (2\pi)^{\frac{1}{2}}$$

$$\int_{-\infty}^{\infty} e^{-Ax^2}\,dx = \left(\frac{\pi}{A}\right)^{\frac{1}{2}} \qquad \int_{0}^{\infty} xe^{-x^2}\,dx = \tfrac{1}{2} \qquad \int_{0}^{\infty} x^2e^{-x^2}\,dx = \frac{\pi^{\frac{1}{2}}}{4}$$

$$\frac{d}{dx}e^{-\pi x^2} = -2\pi xe^{-\pi x^2} \qquad \frac{d^2}{dx^2}e^{-\pi x^2} = -2\pi(1-2\pi x^2)\,e^{-\pi x^2}.$$

Sequences of Gaussian functions play a special role in connection with transforms in the limit. The sequence $\exp(-\pi\tau^2x^2)$ as τ approaches zero is useful for multiplying with functions whose integrals do not converge. The limiting member of the sequence is unity. The sequence $|\tau|^{-1}\exp(-\pi x^2/\tau^2)$ is used for recovering ordinary functions, in cases of impulsive behavior, by convolution. The properties which make the Gaussian function useful in these contexts are that its derivatives are all continuous and that it dies away more rapidly than any power of x; that is,

$$\lim_{x\to\infty} x^n e^{-x^2} = 0$$

for all n.

Figure 4.7 shows these two important sequences; later we emphasize that corresponding members are Fourier transform pairs.

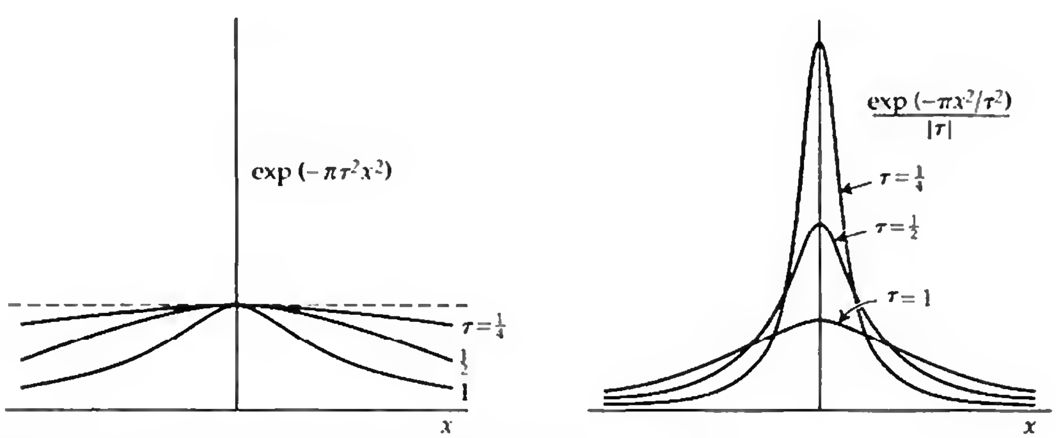


Fig. 4.7 Standard sequences of Gaussian functions.

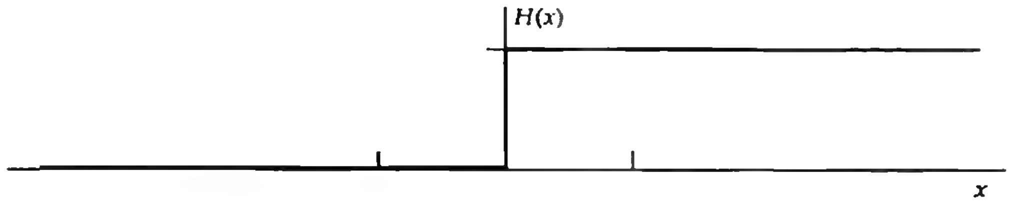


Fig. 4.8 The unit step function.

HEAVISIDE'S UNIT STEP FUNCTION, $H(x)$

An indispensable aid in the representation of simple discontinuities, the unit step function is defined by

$$H(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{2} & x = 0 \\ 1 & x > 0, \end{cases}$$

and is illustrated in Fig. 4.8. It represents voltages which are suddenly switched on or forces which begin to act at a definite time and are constant thereafter. Furthermore, any function with a jump can be decomposed into a continuous function plus a step function suitably displaced. As a simple example of additive use, consider the rectangle function $\Pi(x)$, which has two unit discontinuities, one positive and one negative. If these are removed, nothing remains. Hence $\Pi(x)$ is expressible entirely in terms of step functions as follows (see Fig. 4.9):

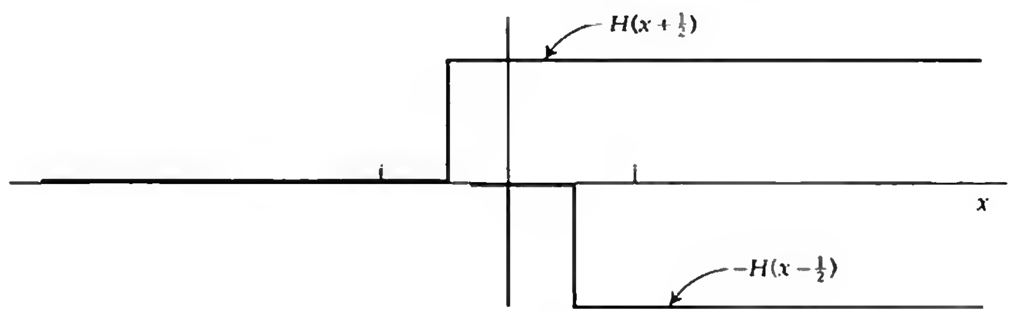


Fig. 4.9 Two functions whose sum is $\Pi(x)$.

Since multiplication of a given function by $H(x)$ reduces it to zero where x is negative but leaves it intact where x is positive, the unit step function provides a convenient way of representing the switching on of simply expressible quantities. For instance, we can represent a sinusoidal quantity which switches on at $t = 0$ by $\sin t\, H(t)$ (see Fig. 4.10). A voltage which has been zero until $t = t_0$ and then jumps to a steady value E is represented by $EH(t - t_0)$ (see Fig. 4.11).

The ramp function $R(x) = xH(x)$ furnishes a further example of notation involving $H(x)$ multiplicatively (see Fig. 4.12). $(F/m)R(t)$ represents the velocity of

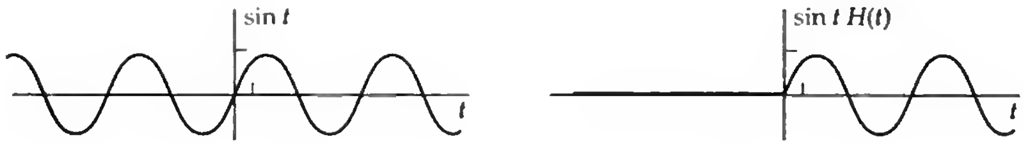


Fig. 4.10 Graphs of sin t and sin t $H(t)$ which exhibit a principal use of the unit step function.

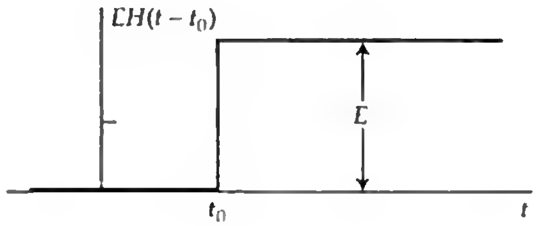


Fig. 4.11 A voltage E which appears at $t = t_0$ represented in step-function notation by $EH(t - t_0)$.

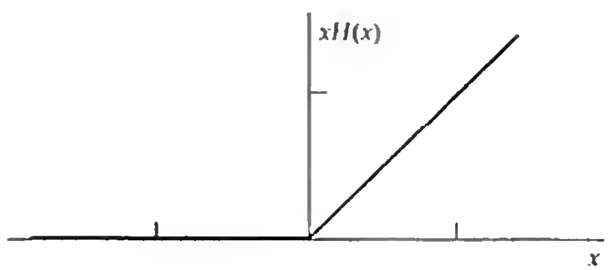


Fig. 4.12 The ramp function $xH(x)$.

a mass m to which a steady force $FH(t)$ has been applied, or the current in a coil of inductance m across which the potential difference is $FH(t)$. It will be seen that $R(x)$ is also the integral of $H(x)$ and, conversely, that $H(x)$ is the derivative of $R(x)$. Thus

$$R(x) = \int_{-\infty}^{x} H(x')\,dx'$$

and

$$R'(x) = H(x).$$

Step-function notation plays a role in simplifying integrals with variable limits of integration by reducing the integrand to zero in the range beyond the original limits. Then constant limits such as $-\infty$ to ∞ or 0 to ∞ can be used. For example, the function $H(x - x')$ is zero where $x' > x$, and therefore $\int_{-\infty}^{x} f(x')\,dx'$ can always be written $\int_{-\infty}^{\infty} f(x')H(x - x')\,dx'$. Thus in the example appearing above we can write

$$R(x) = \int_{-\infty}^{x} H(x')\,dx' = \int_{-\infty}^{\infty} H(x')H(x - x')\,dx'.$$

This last integral reveals the character of $R(x)$ as a convolution integral,

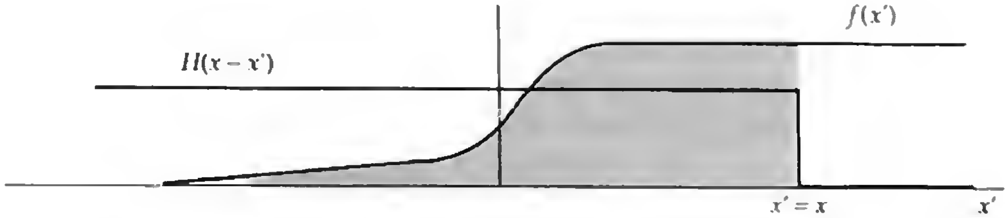


Fig. 4.13 **The shaded area is $\int_{-\infty}^{x} f(x')\,dx'$, or a value of the convolution of $f(x)$ with $H(x)$.**

$$R(x) = H(x) * H(x),$$

and in general we may now note that convolution with $H(x)$ means integration (see Fig. 4.13) provided the integrals exist:

$$H(x) * f(x) = \int_{-\infty}^{\infty} f(x')H(x-x')\,dx' = \int_{-\infty}^{x} f(x')\,dx'$$

or

$$f(x) = \frac{d}{dx}[H(x) * f(x)].$$

Usually, it is not important to define $H(0)$, but for the sake of compatibility with the theory of single-valued functions it is desirable to assign a value, usually $\frac{1}{2}$. Certain internal consistencies are then likely to be observed. For instance, in the equation $R'(x) = H(x)$ it is clear that $R'(0+) = H(0+) = 1$ and that $R'(0-) = H(0-) = 0$, and if we deem $R'(0)$ to mean $\lim_{\Delta x \to 0} [R(x + \frac{1}{2}\Delta x) - R(x - \frac{1}{2}\Delta x)]/\Delta x$, we find $R'(0) = \frac{1}{2}$. Furthermore, the Fourier integral, when it converges at a point of discontinuity, gives the midvalue.

However, there is no obligation to take $H(0) = \frac{1}{2}$; it is not uncommonly taken as zero. This is a natural consequence of a point of view according to which (see Fig. 4.14)

$$\hat{H}(x) = \lim_{\tau \to 0} [(1 - e^{-x/\tau})H(x)].$$

Then $\hat{H}(0) = 0$. The circumflex indicates that there is a slight difference from the definition already decided on.

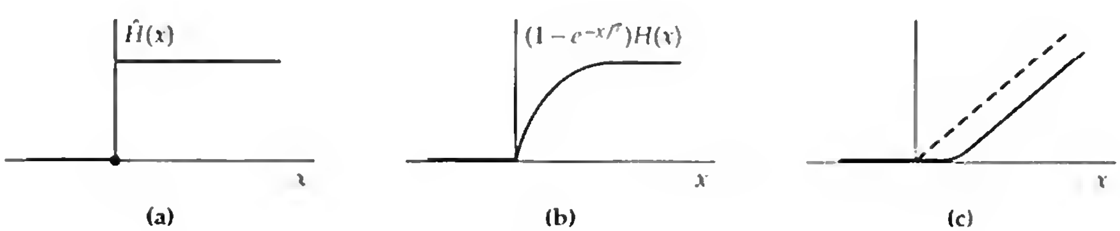


Fig. 4.14 **(a) The function $\hat{H}(x)$; (b) an approximation to $H(x)$; (c) the integral of (b).**

Thus
$$\hat{H}(x) = \begin{cases} 1 & x > 0 \\ 0 & x \le 0 \end{cases}$$

and
$$H(x) = \hat{H}(x) + \tfrac{1}{2}\delta^0(x)$$

where
$$\delta^0(x) = \begin{cases} 0 & x \ne 0 \\ 1 & x = 0. \end{cases}$$

The function $\delta^0(x)$ is one of a class of *null functions* referred to below. No discrepancy need arise from the relation $R'(x) = \hat{H}(x)$, since the integral of $\hat{H}(x)$, regarded as

$$\lim_{\tau\to 0} \int_{-\infty}^{x} (1 - e^{-u/\tau}) H(u)\, du,$$

is certainly $R(x)$; and $R'(0)$, regarded as the limiting slope at $x = 0$ of approximations such as those shown in Fig. 4.14*c*, is certainly equal to $\hat{H}(0)$, namely, 0.

It is sometimes useful to have continuous approximations to $H(x)$. The following examples all approach $H(x)$ as a limit for all x as $\tau \to 0$.

$$\tfrac{1}{2} + \frac{1}{\pi}\arctan\frac{x}{\tau}$$

$$\tfrac{1}{2}\operatorname{erfc}\left(-\frac{x}{\tau}\right) = \frac{1}{\pi^{\frac{1}{2}}}\int_{-x}^{\infty} \tau^{-1} e^{-u^2/\tau^2}\, du$$

$$\tfrac{1}{2} + \frac{1}{\pi}\operatorname{Si}\frac{\pi x}{\tau} = \int_{-\infty}^{x} \tau^{-1} \operatorname{sinc}\frac{u}{\tau}\, du$$

$$\int_{-\infty}^{x} \tau^{-1}\Pi\left(\frac{u}{\tau}\right) du$$

$$\tfrac{1}{2} + \begin{cases} \tfrac{1}{2}(1 - e^{-x/\tau}) & x > 0. \\ -\tfrac{1}{2}(1 - e^{x/\tau}) & x < 0. \end{cases}$$

An example which approaches $H(x)$ as $\tau \to 0$ for all x except $x = 0$ is

$$f(x,\tau) = \begin{cases} 0 & x < -\tau \\ 1 - e^{-(x+\tau)/\tau} & x > -\tau. \end{cases}$$

In this case

$$\lim_{\tau\to 0} f(x,\tau) = H(x) + (\tfrac{1}{2} - e^{-1})\, \delta^0(x),$$

since $f(0,\tau) = 1 - e^{-1}$ for all τ. A further example which approaches $H(x)$ as $\tau \to 0$ is

$$\int_{-\infty}^{x} \tau^{-1}\Lambda\left(\frac{u - \frac{1}{2}\tau}{\tau}\right) du.$$

The difference between $H(x)$ and any version such that $H(0) \ne \tfrac{1}{2}$ is a null function whose integral is always zero. If it were necessary to make physical observations of a quantity varying as $H(x)$ or $\hat{H}(x)$, with the finite resolving power to

which physical observations are limited, it would not be possible to distinguish between the mathematically distinct alternatives, since the weighted means over nonzero intervals, which are the only quantities measurable, would be unaffected by the presence or absence of null functions. For physical applications of $H(x)$ it is therefore perhaps more graceful not to mention $H(0)$.

THE SIGN FUNCTION, sgn x

The function sgn x (pronounced signum x) is equal to $+1$ or -1, according to the sign of x (see Fig. 4.15). Thus

$$\operatorname{sgn} x = \begin{cases} -1 & x < 0 \\ 1 & x > 0. \end{cases}$$

Clearly it differs little from the step function $H(x)$ and has most of its properties. It has a positive discontinuity of 2. The relation to $H(x)$ is

$$\operatorname{sgn} x = 2H(x) - 1.$$

However, $\lim_{A\to\infty} \int_{-A}^{A} \operatorname{sgn} x \, dx = 0$, whereas $\int_{-\infty}^{\infty} H(x)$ does not exist. Furthermore, we may note that sgn x, unlike $H(x)$, is an odd function, and this property distinguishes it sharply from $H(x)$, which has both even and odd parts.

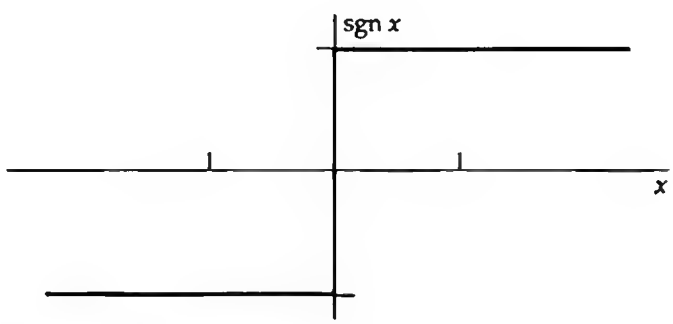


Fig. 4.15 The odd function sgn x.

THE FILTERING OR INTERPOLATING FUNCTION, sinc x

We define

$$\operatorname{sinc} x = \frac{\sin \pi x}{\pi x},$$

a function with the properties that

$$\operatorname{sinc} 0 = 1,$$

$$\operatorname{sinc} n = 0 \qquad n = \text{nonzero integer}$$

$$\int_{-\infty}^{\infty} \operatorname{sinc} x \, dx = 1.$$

This function will appear so frequently, in many different connections, that it is convenient to have a special symbol for it, especially in some agreed normalized form. In the form chosen here, the central ordinate is unity and the total area under the curve is unity (see Fig. 4.16*a*); the word "sinc" appears in Woodward (1953) and has achieved some currency. A table of the sinc function appears on page 508.

The unique properties of sinc x go back to its spectral character: it contains components of all frequencies up to a certain limit and none beyond. Furthermore, the spectrum is flat up to the cutoff frequency. By our choice of notation, sinc x and $\Pi(s)$ are a Fourier transform pair; the cutoff frequency of sinc x is thus 0.5 (cycles per unit of x).

When sinc x enters into convolution it performs ideal low-pass filtering; that is, it removes all components above its cutoff and leaves all below unaltered, and under certain special circumstances discussed later under the sampling theorem, it performs an important kind of interpolation.

In terms of the widely tabulated sine integral Si x (shown in Fig. 4.16*c*), where

$$\operatorname{Si} x = \int_0^x \frac{\sin u}{u} \, du,$$

we have the relations

$$\int_0^x \operatorname{sinc} u \, du = \frac{\operatorname{Si}(\pi x)}{\pi}$$

$$\operatorname{sinc} x = \frac{d}{dx} \frac{\operatorname{Si}(\pi x)}{\pi},$$

and for the integral of sinc x (see Fig. 4.16*b*) we have

$$H(x) * \operatorname{sinc} x = \int_{-\infty}^{x} \operatorname{sinc} u \, du = \frac{1}{2} + \frac{\operatorname{Si}(\pi x)}{\pi}.$$

Another frequently needed function, tabulated on page 508, is the square of sinc x (see Fig. 4.17):

$$\operatorname{sinc}^2 x = \left(\frac{\sin \pi x}{\pi x} \right)^2.$$

This function represents the power radiation pattern of a uniformly excited antenna, or the intensity of light in the Fraunhofer diffraction pattern of a slit. Naturally, it shares with sinc x the property of having a cutoff spectrum, since squar-

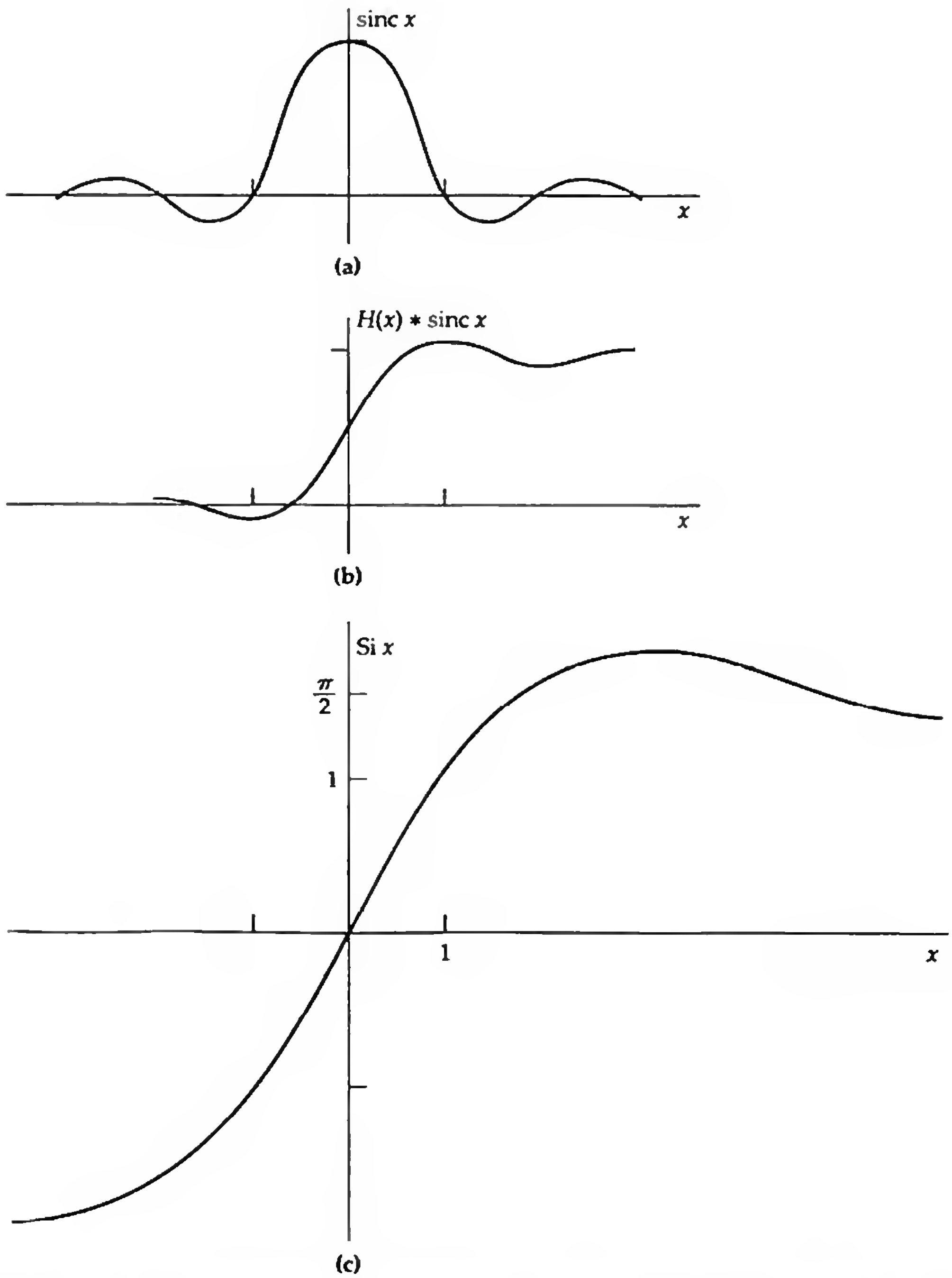


Fig. 4.16 **(a) The filtering or interpolating function sinc *x*; (b) its integral; (c) the sine integral Si *x*.**

ing cannot generate frequencies higher than the sum-frequency of any pair of sinusoidal constituents. A little quantitative thought along the lines of this appeal to physical principles would soon reveal that the Fourier transform of $\text{sinc}^2 x$ is $\Lambda(s)$. The cutoff frequency is one cycle per unit of x.

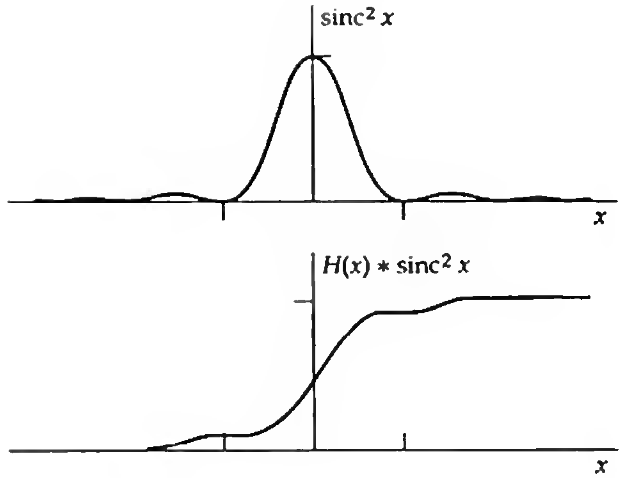


Fig. 4.17 The square of sinc x and its integral.

Among the properties of $\operatorname{sinc}^2 x$ are the following:

$$\operatorname{sinc}^2 0 = 1$$

$$\operatorname{sinc}^2 n = 0 \qquad n = \text{nonzero integer}$$

$$\int_{-\infty}^{\infty} \operatorname{sinc}^2 x \, dx = 1.$$

In two dimensions a function analogous to sinc x is

$$\operatorname{jinc} r = \frac{J_1(\pi r)}{2r},$$

which has unit volume, a central value of $\pi/4$, and a two-dimensional Fourier transform $\Pi(s)$. Another generalization to two dimensions, which has analogous filtering and interpolating properties, is

$$\operatorname{sinc} x \operatorname{sinc} y.$$

The two-dimensional Fourier transform of this function is $\Pi(u)\Pi(v)$.

PICTORIAL REPRESENTATION

Certain conventions in graphical representations will be adopted for the purpose of clarity. For example, theorems relating to Fourier transforms will be illustrated, where possible, by examples which are real. In these diagrams the points where the abscissa and ordinate are equal to unity are marked if it is appropriate to do so.

In representing purely imaginary quantities a dashed line is always used; hence complex quantities may be shown unambiguously and clearly by their real and imaginary parts (see Fig. 4.18).

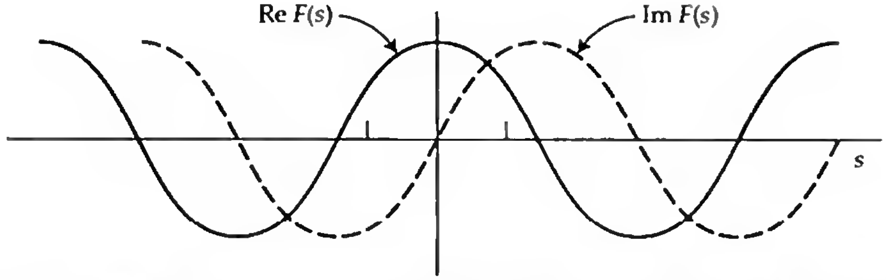


Fig. 4.18 **Representation of a complex function $F(s)$ by its real and imaginary parts.**

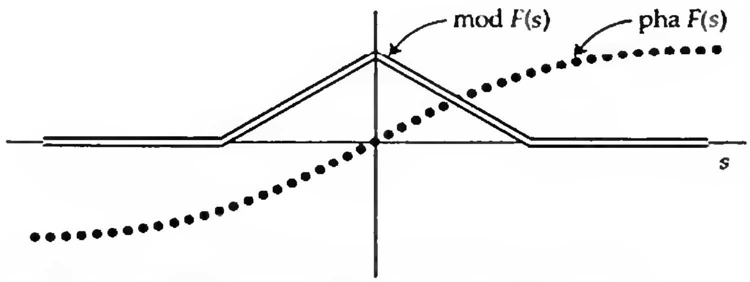


Fig. 4.19 **Complex function shown in modulus and phase.**

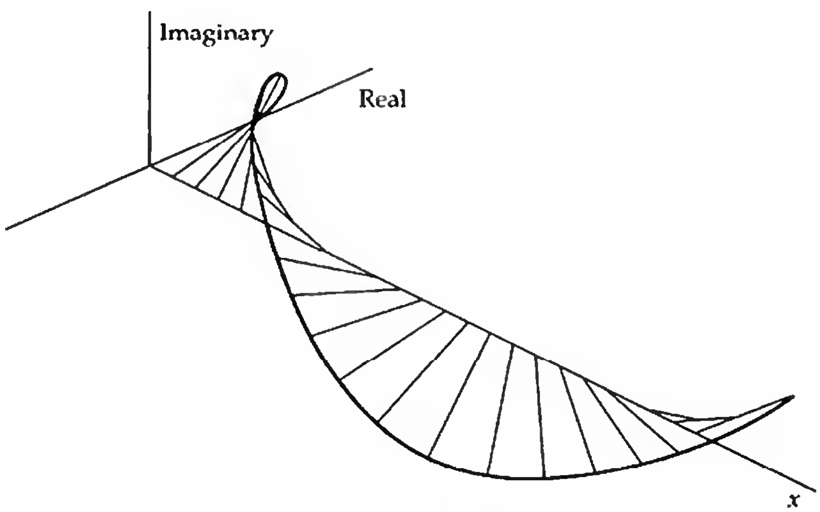


Fig. 4.20 **Complex function in three dimensions.**

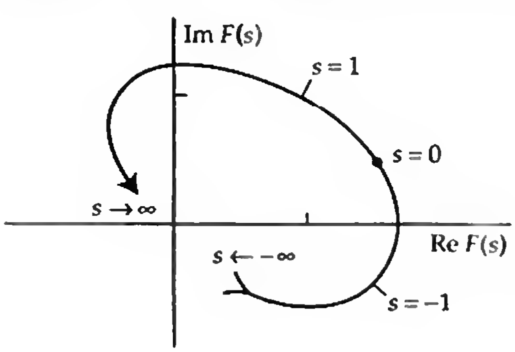


Fig. 4.21 **Values of $F(s)$ on the complex plane of $F(s)$.**

TABLE 4.1
Special symbols

Function	Notation	
Rectangle	$\Pi(x) = \begin{cases} 1 & \lvert x\rvert < \frac{1}{2} \\ 0 & \lvert x\rvert > \frac{1}{2} \end{cases}$	
Triangle	$\Lambda(x) = \begin{cases} 1 - \lvert x\rvert & \lvert x\rvert < 1 \\ 0 & \lvert x\rvert > 1 \end{cases}$	
Heaviside unit step	$H(x) = \begin{cases} 0 & x > 0 \\ 0 & x < 0 \end{cases}$	
Sign (signum)	$\operatorname{sgn} x = \begin{cases} 1 & x > 0 \\ -1 & x < 0 \end{cases}$	
Impulse symbol*	$\delta(x)$	
Sampling or replicating symbol*	$\mathrm{III}(x) = \sum_{-\infty}^{\infty} \delta(x - n)$	
Even impulse pair	$\mathrm{II}(x) = \frac{1}{2}\delta(x + \frac{1}{2}) + \frac{1}{2}\delta(x - \frac{1}{2})$	
Odd impulse pair	$\mathrm{I_I}(x) = \frac{1}{2}\delta(x + \frac{1}{2}) - \frac{1}{2}\delta(x - \frac{1}{2})$	
Filtering or interpolating	$\operatorname{sinc} x = \dfrac{\sin \pi x}{\pi x}$	
Jinc function	$\operatorname{jinc} x = \dfrac{J_1(\pi x)}{2x}$	
Asterisk notation for convolution	$f(x) * g(x) \triangleq \int_{-\infty}^{\infty} f(u) g(x - u)\, du$	(p. 24)
Asterisk notation for serial products	$\{f_i\} * \{g_i\} \triangleq \sum_j f_j g_{i-j}$	(p. 31)
Pentagram notation	$f(x) \star g(x) \triangleq \int_{-\infty}^{\infty} f(u) g(x + u)\, du$	(p. 40)
Various two-dimensional functions	${}^2\Pi(x,y) = \Pi(x)\Pi(y)$ ${}^2\delta(x,y) = \delta(x)\,\delta(y)$ ${}^2\mathrm{III}(x,y) = \mathrm{III}(x)\mathrm{III}(y)$ ${}^2\operatorname{sinc}(x,y) = \operatorname{sinc} x \operatorname{sinc} y$	

*See Chapter 5.

Sometimes a complex function may be shown to advantage by plotting its modulus and phase, as in Fig. 4.19. Another method is to show a three-dimensional diagram (see Fig. 4.20), and a third method is to show a locus on the complex plane (see Fig. 4.21).

SUMMARY OF SPECIAL SYMBOLS

A small number of special symbols which are used extensively throughout this work are summarized in Table 4.1 and in Fig. 4.22 for reference.

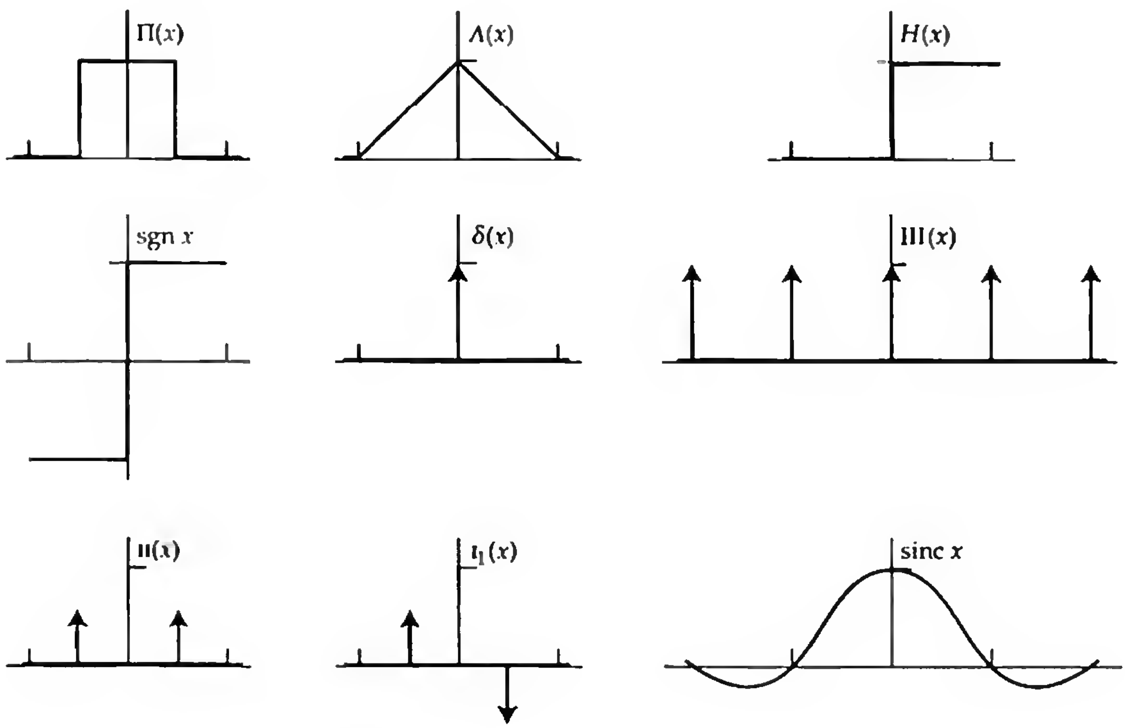


Fig. 4.22 **Summary of special symbols.**

BIBLIOGRAPHY

Woodward, P. M.: "Probability and Information Theory with Applications to Radar," McGraw-Hill Book Company, New York, 1953.

PROBLEMS

1. Show that

$$H(ax+b) = \begin{cases} H\left(x + \dfrac{b}{a}\right) & a > 0 \\ H\left(-x - \dfrac{b}{a}\right) & a < 0, \end{cases}$$

and hence that

$$H(ax+b) = H\left(x + \frac{b}{a}\right)H(a) + H\left(-x - \frac{b}{a}\right)H(-a).$$

2. Discuss the function $\frac{1}{2}[1 + x/(x^2)^{\frac{1}{2}}]$ used by Cauchy.

3. Show that the operation $H(x) *$ is an integrating operation in the sense that

$$H(x) * [f(x)H(x)] = \int_0^x f(x)\,dx.$$

4. Calculate $(d/dx)[\Pi(x) * H(x)]$ and prove that $(d/dx)[f(x) * H(x)] = f(x)$.

5. By evaluating the integral, prove that $\operatorname{sinc} x * \operatorname{sinc} x = \operatorname{sinc} x$.

6. Prove that $\operatorname{sinc} x * J_0(\pi x) = J_0(\pi x)$.

7. Prove that $4 \operatorname{sinc} 4x * \sin x = \sin x$.▷

8. Show that

$$\begin{aligned} \Pi(x) &= H(x + \tfrac{1}{2}) - H(x - \tfrac{1}{2}) \\ &= H(\tfrac{1}{2} + x) + H(\tfrac{1}{2} - x) - 1 \\ &= H(\tfrac{1}{4} - x^2) \\ &= \tfrac{1}{2}[\operatorname{sgn}(x + \tfrac{1}{2}) - \operatorname{sgn}(x - \tfrac{1}{2})] \end{aligned}$$

and that $\Pi(x^2) = \Pi(x/2^{\frac{1}{2}})$.

9. Show that

$$\begin{aligned} \Lambda(x) &= \Pi(x) * \Pi(x) \\ &= \Pi(x) * H(x + \tfrac{1}{2}) - \Pi(x) * H(x - \tfrac{1}{2}). \end{aligned}$$

10. Experiment with the equation $f[f(x)] = f(x)$ and note that $f(x) = \operatorname{sgn} x$ is a solution. Find other solutions and attempt to write down the general solution compactly with the aid of step-function notation.

11. Show that $\operatorname{erf} x = 2\Phi(2^{\frac{1}{2}}\sigma x) - 1$, where

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sigma(2\pi)^{\frac{1}{2}}} e^{-u^2/2\sigma^2}\,du.$$

12. Show that the first derivative of $\Lambda(x)$ is given by

$$\Lambda'(x) = -\Pi\left(\frac{x}{2}\right)\operatorname{sgn} x$$

and calculate the second derivative. ▷

13. In abbreviated notation the relation of $\Lambda(x)$ to $\Pi(x)$ could be written $\Lambda = \Pi * \Pi$ or $\Lambda = \Pi^{*2}$. Show that

$$\Pi^{*2} = \Pi * \Lambda = \tfrac{1}{2}(x + 1\tfrac{1}{2})^2\Pi(x + 1) + (\tfrac{3}{4} - x^2)\Pi(x) + \tfrac{1}{2}(x - 1\tfrac{1}{2})^2\Pi(x - 1).$$

Show also that

$$\Pi^{*2} = \tfrac{1}{2}(x + 1\tfrac{1}{2})^2 H(x + 1\tfrac{1}{2}) - \tfrac{3}{2}(x + \tfrac{1}{2})^2 H(x + \tfrac{1}{2}) + \tfrac{3}{2}(x - \tfrac{1}{2})^2 H(x - \tfrac{1}{2}) - \tfrac{1}{2}(x - 1\tfrac{1}{2})^2 H(x - 1\tfrac{1}{2}).$$

14. Examine the derivatives of Π^{*3} at $x = \frac{1}{2}\pm$ and $x = 1\frac{1}{2}\pm$ and reach some conclusion about the continuity of slope and curvature.

15. Show that $(d/dx)|x| = \operatorname{sgn} x$ and that $(d/dx)\operatorname{sgn} x = 2\delta(x)$. Comment on the fact that

$$\frac{d^2|x|}{dx^2} = \frac{d^2}{dx^2}[2xH(x)] = 2\delta(x). \ ▷$$

16. Notation. Prove that

$$H(x) * H(x + 1) - H(x) * H(x - 1) = \Lambda(x),$$

or, if the RHS is not correct, derive the correct expression.

17. Notation. Prove that

$$\frac{d}{dx}\{[\Lambda(x)]^2\} = -\Lambda(x)\operatorname{sgn} x,$$

or, if the RHS is not correct, derive the correct expression.

18. Derivatives of the sinc function. Show that

$$\operatorname{sinc}'(x) = \frac{\cos \pi x}{x} - \frac{\sin \pi x}{\pi x^2},$$

$$\operatorname{sinc}''(x) = -\frac{2}{x^2}\cos \pi x + \frac{2x + \pi^2 x^3}{\pi x^4}\sin \pi x.$$

19. Integrals of the jinc function. When the jinc function was introduced it was mentioned that $\int_0^\infty \operatorname{jinc} r\, 2\pi r\, dr = 1$, i.e., regarded as a function of radius in two dimensions, the volume under the jinc function is unity. Confirm that $\int_{-\infty}^{\infty} \operatorname{jinc} x\, dx$ is also equal to unity. ▷

5

The Impulse Symbol

It is convenient to have notation for intense unit-area pulses so brief that measuring equipment of a given resolving power is unable to distinguish between them and even briefer pulses. This concept is covered in mechanics by the term "impulse." The important attribute of an impulse is its integral; the precise details of its form are unimportant. The idea has been current for a century or more in mathematical circles. For historical examples from the writings of Hermite, Cauchy, Poisson, Kirchhoff, Helmholtz, Kelvin, and Heaviside, see van der Pol and Bremmer (1955). The notation $\delta(x)$, first used by G. Kirchhoff, was subsequently introduced into quantum mechanics in 1927 by Dirac (1947), and is now in general use. The underlying concept permeates physics. Point masses, point charges, point sources, concentrated forces, line sources, surface charges, and the like are familiar and accepted entities in physics. Of course, these things do not exist. Their conceptual value stems from the fact that the impulse response—the effect associated with the impulse (point mass, point charge, and the like)—may be indistinguishable, given measuring equipment of specified resolving power, from the response due to a physically realizable pulse. It is then a convenience to have a name for pulses which are so brief and intense that making them any briefer and more intense does not matter.

We have in mind an infinitely brief or concentrated, infinitely strong unit-area impulse and therefore wish to write

$$\delta(x) = 0 \qquad x \neq 0$$

and

$$\int_{-\infty}^{\infty} \delta(x)\, dx = 1.$$

However, the impulse symbol[1] $\delta(x)$ does not represent a function in the sense in which the word is used in analysis (to stress this fact Dirac coined the term "im-

[1] I use the word "symbol" systematically to signpost the entities that are not functions; one may also use the term "generalized function," introduced in 1953 by G. Temple. However, the distinction between a function $f(x)$ and the entity $\delta(x)$ is glossed over in casual writing, where "delta function" is in common use for the understandable reason that $\delta(x)$ handles like a function.

proper function"), and the above integral is not a meaningful quantity until some *convention* for interpreting it is declared. Here we use it to mean

$$\lim_{\tau \to 0} \int_{-\infty}^{\infty} \tau^{-1} \Pi\left(\frac{x}{\tau}\right) dx.$$

The function $\tau^{-1}\Pi(x/\tau)$ is a rectangle function of height τ^{-1} and base τ and has unit area; as τ tends to zero a sequence of unit-area pulses of ever increasing height is generated. The limit of the integral is, of course, equal to unity. In other words, to interpret expressions containing the impulse symbol, we fall back on certain sequences of finite unit-area pulses of brief but nonzero duration, and of some particular shape. We perform the operations indicated, such as integration, differentiation, multiplication, and then discuss limits as the duration approaches zero. There is some convenience mathematically in taking the pulse shape to be always Gaussian; obviously, one has to prepare for possible awkwardness in retaining $\Pi(x)$ as a choice, where differentiation is involved. However, the essence of the physics is that the pulse shape should not matter, and we therefore proceed under the expectation that the choice of pulse shape will remain at our disposal.

We adapt our approach to each case as it arises. Later we give a systematic presentation of the theory of generalized functions, a rigorously developed exposition of the sequence idea in tidied-up mathematical form.

The need to broaden Fourier transform theory was mentioned earlier in connection with functions, such as periodic functions, which do not possess Fourier transforms. The term "transform pair in the limit" was introduced to describe cases where one or both members of the transform pair are generalized functions. All these cases can be expressed with the aid of the impulse symbol and its derivatives, $\delta(x)$, $\delta'(x)$, and so on, which thus furnish the notation for the broadened theory.

The convenience of the impulse symbol lies in its reserve over detail. As a specific example of the relevance of this feature to physical systems, consider an electrical network, say a low-pass filter. An applied pulse of voltage produces a certain transient response, and it is readily observable, as the applied pulses are made briefer and briefer, that the response settles down to a definite form. It is also observable that the form of the response is then independent of the input pulse shape, be it rectangular, triangular, or even a pair of pulses. This happens because the high-frequency components, which distinguish the different applied pulses, produce negligible response. The network is thus characterized by a certain definite, readily observable form of response, which can be elicited by a multiplicity of applied waveforms, the details of which are irrelevant; it is necessary only that they be brief enough. Since the response may be scrutinized with an oscilloscope of the highest precision and time resolution, we must, of course, be prepared to keep the applied pulse duration shorter than the minimum set by the quality of the measuring instrument. The impulse symbol enables us to make abbreviated statements about arbitrarily shaped indefinitely brief pulses.

An intimate relationship between the impulse symbol and the unit step function follows from the property that $\int_{-\infty}^{x} \delta(x')\, dx'$ is unity if x is positive but zero

if x is negative. Hence

$$\int_{-\infty}^{x} \delta(x')\,dx' = H(x).$$

This equation furnishes an opportunity to illustrate the interpretation of an expression containing the impulse symbol. First we replace $\delta(x)$ by the pulse sequence $\tau^{-1}\Pi(x/\tau)$ and contemplate the sequence of integrals

$$\int_{-\infty}^{x} \tau^{-1}\Pi\left(\frac{x'}{\tau}\right)dx'.$$

As long as τ is not zero or infinite, each such integral is a function of x that may be described as a ramp-step function, as illustrated in Fig. 5.1. Now fix x, and consider the limit of the sequence of values of the integral generated as τ approaches zero. We see that if we have fixed on a positive (negative) x, then the limit of the integral will be unity (zero). Therefore, in accordance with the definition of $H(x)$ we can write

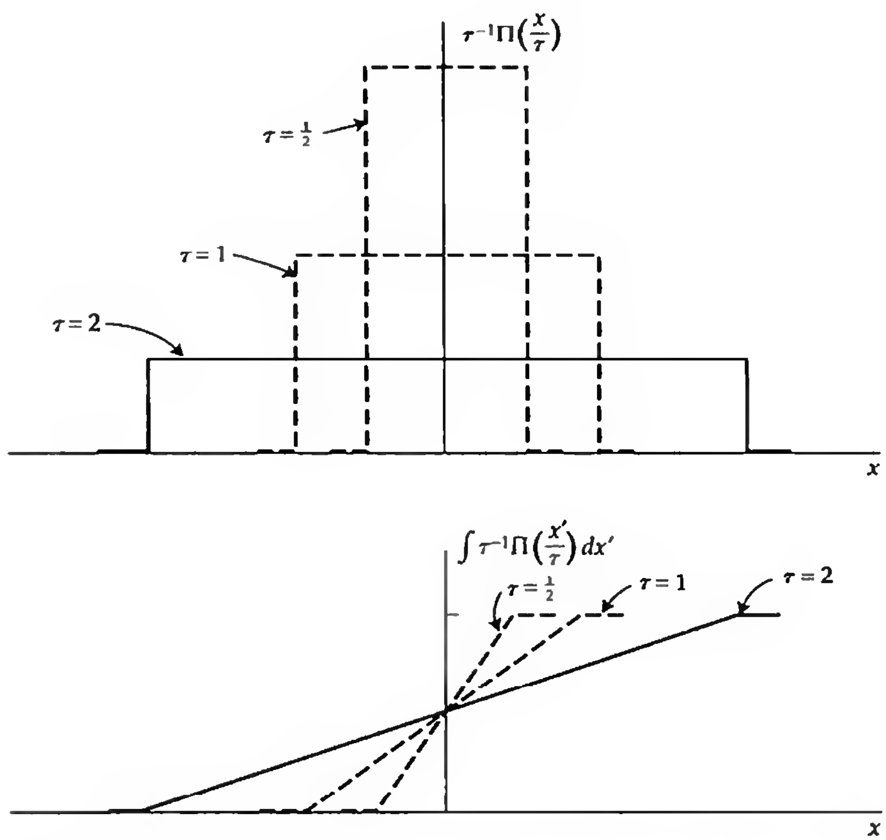


Fig. 5.1 A rectangular pulse sequence and the sequence of ramp-step functions obtained by integration.

$$\lim_{\tau \to 0} \int_{-\infty}^{x} \tau^{-1} \Pi\left(\frac{x'}{\tau}\right) dx' = H(x).$$

The equation

$$\int_{-\infty}^{x} \delta(x')\, dx' = H(x)$$

is shorthand for this.

Since under ordinary circumstances

$$\int_{-\infty}^{x} f(x')\, dx' = g(x)$$

implies that

$$f(x) = g'(x),$$

one writes by analogy

$$\delta(x) = \frac{d}{dx} H(x)$$

and states that "the derivative of the unit step function is the impulse symbol." Since the unit step function does not in fact possess a derivative at the origin, this statement must be interpreted as shorthand for "the derivatives of a sequence of differentiable functions that approach $H(x)$ as a limit constitute a suitable defining sequence for $\delta(x)$." The ramp-step functions of Fig. 5.1 are differentiable and approach $H(x)$. Since the amount of the step is in all cases unity, the area under each derivative function is unity, thus qualifying the sequence of derivative functions as a suitable sequence to define a unit impulse.

It is not *necessary* to deal in terms of sequences of *rectangle* functions to discuss impulses. Representations of $\delta(x)$ in terms of various pulse shapes include the following sequences, generated as τ approaches zero (through positive values).

The sequence

$$\tau^{-1} \Pi\left(\frac{x - \frac{1}{2}\tau}{\tau}\right)$$

is composed of rectangle functions, all having their left-hand edge at the origin; it is often considered in the analysis of circuits where transients cannot exist prior to $t = 0$, the instant at which some switch is thrown. If one uses this sequence instead of the centered sequence, the results will not necessarily be the same. This is discussed below in connection with the sifting property. The sequence

$$\tau^{-1} e^{-\pi x^2/\tau^2}$$

of Gaussian profiles, which has been mentioned previously, has the convenient property that derivatives of all orders exist. On the other hand, these profiles lack the convenience of being nonzero over only a finite range of x. The sequence

$$\tau^{-1}\Lambda\left(\frac{x}{\tau}\right)$$

of triangle functions is useful for discussing situations where first derivatives are needed, since the profiles are continuous. In addition, they have an advantage in being zero outside the interval in which $|x| < \tau$. The sequence

$$\tau^{-1}\,\text{sinc}\,\frac{x}{\tau}$$

has the curious property of not dying out to zero where $x \neq 0$; at any value of x not equal to zero the value oscillates without diminishing as $\tau \to 0$. The sequence serves perfectly well to define $\delta(x)$ for a reason that is given below in connection with the sifting property. The resonance profiles

$$\frac{\tau}{\pi(x^2 + \tau^2)}$$

decay rather slowly with increasing x. A product with an arbitrary function may well have infinite area. To eliminate this possibility completely, one is led to contemplate sequences that are zero outside a finite range, thus obtaining freedom to accept products with functions having any kind of asymptotic behavior. In addition, one would like to have derivatives of all orders exist. The expression

$$\tau^{-1}\exp\left[\frac{-1}{1-(x/\tau)^2}\right]\Pi\left(\frac{x}{2\tau}\right)$$

is a specific example of a sequence of profiles, each of which is zero outside the interval in which $|x| < \tau$, and each of which possesses derivatives of all orders. To see that it is possible for a function to descend to zero with zero slope, zero curvature, and all higher derivatives zero, differentiate the function

$$e^{-1/x}H(x),$$

and evaluate the derivatives at $x = 0$. There may seem to be a clash with the Maclaurin formula, according to which

$$f(x) = f(0) + xf'(0) + \frac{x^2}{2}f''(0) + \dots + \frac{x^n}{n!}f^{(n)}(0) + \text{remainder}.$$

In the case of many functions familiar from analysis, the remainder term can be shown to vanish as $n \to \infty$. In the case we have chosen here, however, the first $n + 1$ terms vanish and the remainder term contains the whole value of the function.

THE SIFTING PROPERTY

Following our rule for interpreting expressions containing the impulse symbol, we may try to assign a meaning to

$$\int_{-\infty}^{\infty} \delta(x) f(x)\, dx.$$

Thus we substitute the sequence $\tau^{-1}\Pi(x/\tau)$ for $\delta(x)$, perform the multiplication and integration, and finally take the limit of the integral as $\tau \to 0$:

$$\lim_{\tau \to 0} \int_{-\infty}^{\infty} \tau^{-1}\Pi\left(\frac{x}{\tau}\right) f(x)\, dx.$$

In Fig. 5.2 the integrand is indicated in broken outline. Its area is τ^{-1} times the shaded area. The shaded area, whose width is τ and whose average height is approximately $f(0)$, amounts to approximately $\tau f(0)$. Hence the area under the integrand approaches $f(0)$ as τ approaches zero. Thus we write

$$\int_{-\infty}^{\infty} \delta(x) f(x)\, dx = f(0)$$

and refer to this statement as the sifting property of the impulse symbol, since the operation on $f(x)$ indicated on the left-hand side sifts out a single value of $f(x)$. It will be seen that it is immaterial what sort of pulse is incorporated in the integrand, and this fact is the essence of the utility of $\delta(x)$. It just stands for a unit pulse whose duration is much smaller than any interval of interest, and consequently whose pulse *shape* means nothing to us; only its integral counts. It will be represented on graphs (see Fig. 5.3) as a spike of unit height, and impulses in general will be shown as spikes of height equal to their integral.

It is clear that we can also write

$$\int_{-\infty}^{\infty} \delta(x - a) f(x)\, dx = f(a)$$

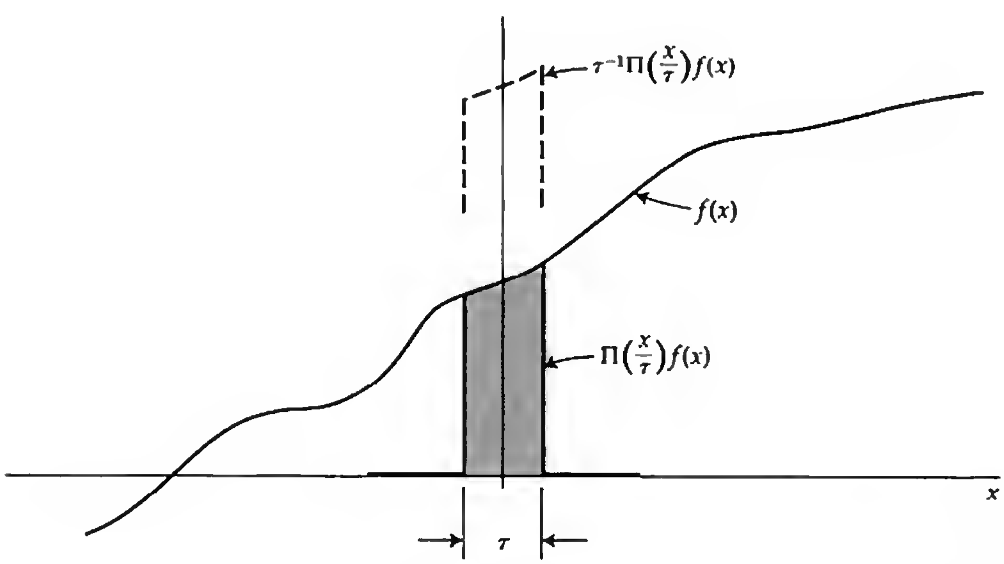


Fig. 5.2 **Explaining the sifting property. The shaded area is approximately $\tau f(0)$.**

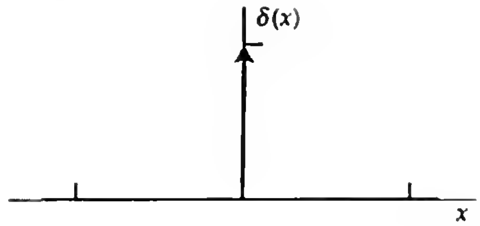


Fig. 5.3 **Graphical representation of the impulse symbol $\delta(x)$ as a spike of unit height.**

and
$$\int_{-\infty}^{\infty} \delta(x)f(x-a)\,dx = f(-a).$$

The resemblance to the convolution integral can be emphasized by writing

$$\int_{-\infty}^{\infty} \delta(x')f(x-x')\,dx' = \int_{-\infty}^{\infty} \delta(x-x')f(x')\,dx' = f(x)$$

or, in asterisk notation,

$$\delta(x) * f(x) = f(x) * \delta(x) = f(x).$$

If $f(x)$ has a jump at $x = 0$, a little thought devoted to a diagram such as Fig. 5.2 will show that the sifting integral will have a limiting value of $\frac{1}{2}[f(0+) + f(0-)]$. Consequently, it is more general to write

$$\delta(x) * f(x) = \frac{f(x+) + f(x-)}{2}.$$

The expression on the right-hand side differs from $f(x)$ only by a null function, and hence the refinement is ordinarily not important. This does not alter the fact that $\frac{1}{2}[f(x+) + f(x-)]$ can be different in value from $f(x)$.

The asymmetrical sequence $\tau^{-1}\Pi[(x - \frac{1}{2}\tau)/\tau]$ mentioned above will be seen to have the property of sifting out $f(x+)$. At points of discontinuity of $f(x)$, the use of this asymmetrical sequence therefore gives a different result. In transient analysis, where discontinuities at the switching instant $t = 0$ are particularly common, the choice of sequence can thus appear to give different answers. The difference, however, can only be instantaneous.

The impulse symbol has many fascinating properties, most of which can be proved easily. An important one which must be watched carefully in algebraic manipulation is

$$\delta(ax) = \frac{1}{|a|}\,\delta(x);$$

that is, if the scale of x is compressed by a factor a, thus reducing the area of the pulses which previously had unit area, then the strength of the impulse is reduced by the factor $|a|$. The modulus sign allows for the property

$$\delta(-x) = \delta(x).$$

From this it would seem that the impulse symbol has the property of evenness; however, we gave an equation earlier involving a sequence of displaced rectangle functions which were not themselves even (Prob. 19, Chapter 5).

It can easily be shown by considering sequences of pulses that we may write, if $f(x)$ is continuous at $x = 0$,

$$f(x)\,\delta(x) = f(0)\,\delta(x).$$

From the sifting property, putting $f(x) = x$, we have

$$\int_{-\infty}^{\infty} x\,\delta(x)\,dx = 0.$$

One generally writes

$$x\,\delta(x) \equiv 0,$$

although if the prelimit graphs are contemplated, it will be seen that this equation conceals a nonvanishing component reminiscent of the Gibbs phenomenon in Fourier series. Thus it is true that

$$\lim_{\tau\to 0}\left[x\tau^{-1}\Pi\left(\frac{x}{\tau}\right)\right] = 0 \qquad \text{for all } x;$$

nevertheless,

$$\lim_{\tau\to 0}\left[x\tau^{-1}\Pi\left(\frac{x}{\tau}\right)\right]_{\max} = \tfrac{1}{2},$$

and, moreover, the limit of the minimum value is $-\frac{1}{2}$. Consequently, among those functions which are identically zero, $x\delta(x)$ is rather curious, and one has the feeling that if it could be applied to the deflecting electrodes of an oscilloscope, one would see spikes.

THE SAMPLING OR REPLICATING SYMBOL III(x)

Consider an infinite sequence of unit impulses spaced at unit interval as shown in Fig. 5.4. Any reservations that apply to the impulse symbol $\delta(x)$ apply equally

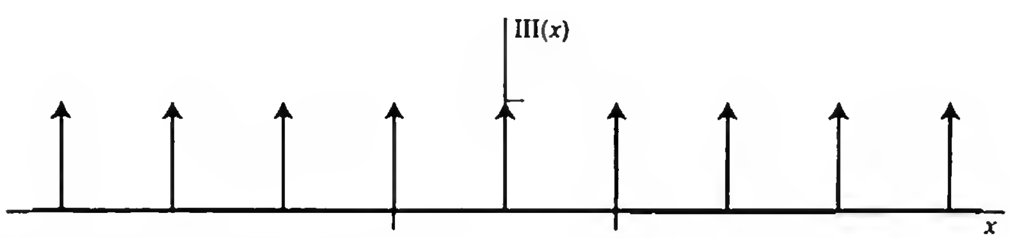


Fig. 5.4 The shah symbol III(x).

in this case; indeed, even more may be needed because we have to deal with an infinite number of infinite discontinuities and a nonconvergent infinite integral. For example, *all* the conditions for existence of a Fourier transform are violated. The conception of an infinite sequence of impulses proves, however, to be extremely useful—and easy to manipulate algebraically.

To describe this conception we introduce the *shah*[2] symbol $\mathrm{III}(x)$ and write

$$\mathrm{III}(x) = \sum_{n=-\infty}^{\infty} \delta(x - n).$$

Various obvious properties may be pointed out:

$$\begin{aligned}
\mathrm{III}(ax) &= \frac{1}{|a|}\sum \delta\left(x - \frac{n}{a}\right)\\
\mathrm{III}(-x) &= \mathrm{III}(x)\\
\mathrm{III}(x+n) &= \mathrm{III}(x) \qquad n \text{ integral}\\
\mathrm{III}(x-\tfrac{1}{2}) &= \mathrm{III}(x+\tfrac{1}{2})\\
\int_{n-\frac{1}{2}}^{n+\frac{1}{2}} \mathrm{III}(x)\,dx &= 1\\
\mathrm{III}(x) &= 0 \qquad x \neq n.
\end{aligned}$$

Evidently, $\mathrm{III}(x)$ is periodic with unit period.

A periodic *sampling* property follows as a generalization of the sifting integral already discussed in connection with the impulse symbol. Thus multiplication of a function $f(x)$ by $\mathrm{III}(x)$ effectively samples it at unit intervals:

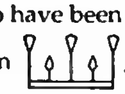
$$\mathrm{III}(x)f(x) = \sum_{n=-\infty}^{\infty} f(n)\,\delta(x - n).$$

The information about $f(x)$ in the intervals between integers where $\mathrm{III}(x) = 0$ is not contained in the product; however, the values of $f(x)$ at integral values of x are preserved (see Fig. 5.5).

The sampling property makes $\mathrm{III}(x)$ a valued tool in the study of a wide variety of subjects (for example, the radiation patterns of antenna arrays, the diffraction patterns of gratings, raster scanning in television and radar, pulse modulation, data sampling, Fourier series, and computing at discrete tabular intervals).

Just as important as the sampling property under multiplication is a *replicating* property exhibited when $\mathrm{III}(x)$ enters into convolution with a function $f(x)$. Thus

---

[2]The symbol Ш is pronounced *shah* after the Cyrillic character Ш, which is said to have been modeled on the Hebrew letter ש (*shin*), which in turn may derive from the Egyptian , a hieroglyph depicting papyrus plants along the Nile.

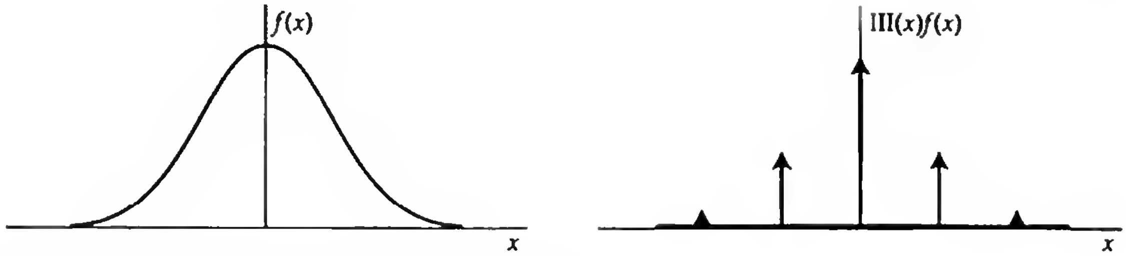


Fig. 5.5 The sampling property of III(x).

$$\text{III}(x) * f(x) = \sum_{n=-\infty}^{\infty} f(x - n);$$

as shown in Fig. 5.6, the function $f(x)$ appears in replica at unit intervals of x *ad infinitum* in both directions. Of course, if $f(x)$ spreads over a base more than one unit wide, there is overlapping.

The III symbol is thus also applicable wherever there are periodic structures. This twofold character is not accidental, but is connected with the fact that III is its own Fourier transform (in the limit), which of course makes it twice as useful as it otherwise would have been.

The self-reciprocal property under the Fourier transformation is derived later.

When a shah function is squeezed in the x-direction by a factor 2, the result being written III($2x$), the impulses are packed twice as closely. But for algebraic consistency the impulses are reduced in strength by the same factor. One can understand this in terms of the sequence of rectangle functions defining each impulse; each rectangle is squeezed to half the width and half the area. Therefore, just as $\delta(ax) = |a|^{-1}\,\delta(x)$ ($|a|$ appears in case $a < 0$), so

$$\text{III}(ax) = \frac{1}{|a|} \sum_n \delta\left(x - \frac{n}{a}\right).$$

This is a known trap for students and may be restated as follows. If III(x) is stretched so that the impulse spacing changes from unity to X, then III(x/X) represents impulses at spacing X but they are no longer of unit strength X. The expression for *unit* impulses spaced X is $X^{-1}\text{III}(x/X)$.

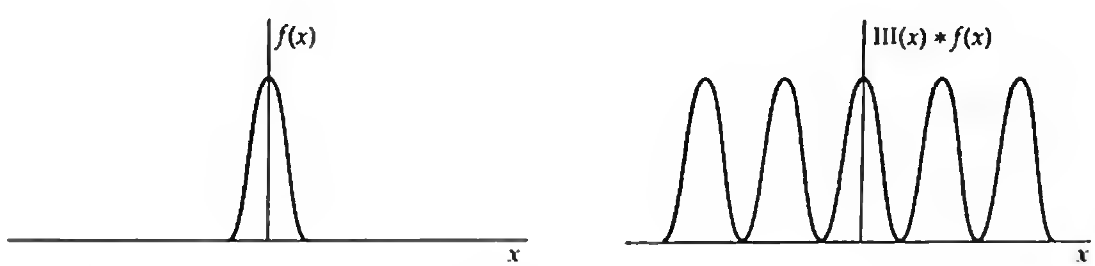


Fig. 5.6 The replicating property of III(x).

THE EVEN AND ODD IMPULSE PAIRS $\mathrm{II}(x)$ and $\mathrm{I}_\mathrm{I}(x)$

Figure 5.7 shows the often-needed impulse-pair symbols defined by

$$\mathrm{II}(x) = \tfrac{1}{2}\delta(x + \tfrac{1}{2}) + \tfrac{1}{2}\delta(x - \tfrac{1}{2}),$$
$$\mathrm{I}_\mathrm{I}(x) = \tfrac{1}{2}\delta(x + \tfrac{1}{2}) - \tfrac{1}{2}\delta(x - \tfrac{1}{2}).$$

The impulse pairs derive importance from their transform relationship to the cosine and sine functions. Thus

$$\mathrm{II}(x) \supset \cos \pi s \qquad \cos \pi x \supset \mathrm{II}(s)$$
$$\mathrm{I}_\mathrm{I}(x) \supset i \sin \pi s \qquad \sin \pi x \supset i\, \mathrm{I}_\mathrm{I}(s).$$

When convolved with a function $f(x)$, the even impulse pair $\mathrm{II}(x)$ has a duplicating property. Thus, as illustrated in Fig. 5.8,

$$\mathrm{II}(x) * f(x) = \tfrac{1}{2}f(x + \tfrac{1}{2}) + \tfrac{1}{2}f(x - \tfrac{1}{2}).$$

There are occasions when $\mathrm{II}(x)$ might better consist of two unit impulses, but as defined it is normalized to unit area; that is,

$$\int_{-\infty}^{\infty} \mathrm{II}(x)\, dx = 1,$$

which has advantages.

If the finite difference of $f(x)$ is defined by

$$\Delta f(x) = f(x + \tfrac{1}{2}) - f(x - \tfrac{1}{2}),$$

then

$$\Delta f(x) = 2\, \mathrm{I}_\mathrm{I}(x) * f(x).$$

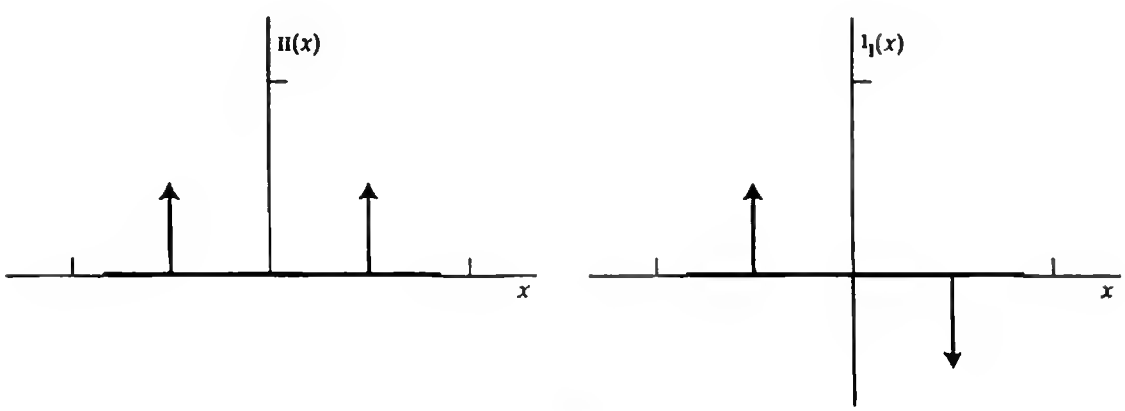


Fig. 5.7 **The even and odd impulse pairs, $\mathrm{II}(x)$ and $\mathrm{I}_\mathrm{I}(x)$.**

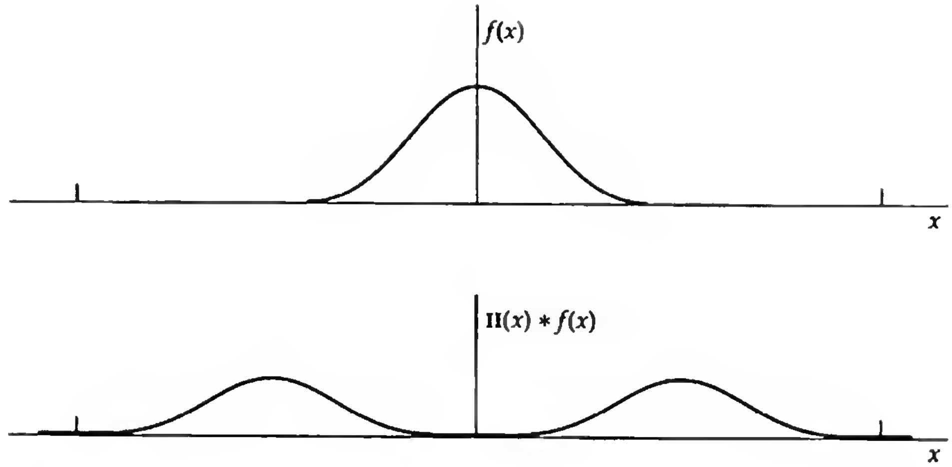


Fig. 5.8 **The convolution of $\mathrm{II}(x)$ with $f(x)$.**

Thus the finite difference operator can be expressed as

$$\Delta \equiv 2\,\mathrm{I_I}\,*.$$

DERIVATIVES OF THE IMPULSE SYMBOL

The first derivative of the impulse symbol is defined symbolically by

$$\delta'(x) = \frac{d}{dx}\,\delta(x).$$

The mental picture that accompanies this conception is the same as that involved in the conception of an infinitesimal dipole in electrostatics, and it is well known that the idea of an infinitesimal dipole is convenient and easy to think with physically. The $\delta'(x)$ notation carries this facility over into mathematical form, but, of course, there are difficulties because we cannot ask a function to go positively infinite just to the left of the origin and negatively infinite just to the right, and to be zero where $|x| > 0$. To cap this we would wish to write $\delta'(0) = 0$.

For rigorous interpretation of statements involving $\delta'(x)$ we may fall back on sequences of pulses such as were invoked in connection with $\delta(x)$, and consider their derivatives (two examples are given in Fig. 5.9). Then statements such as

$$\int_{-\infty}^{\infty} \delta'(x)\,dx = 0$$

are deemed to be shorthand for statements such as

$$\lim_{\tau\to 0} \int_{-\infty}^{\infty} \left[\frac{-2\tau x}{\pi(x^2 + \tau^2)^2}\right] dx = 0,$$

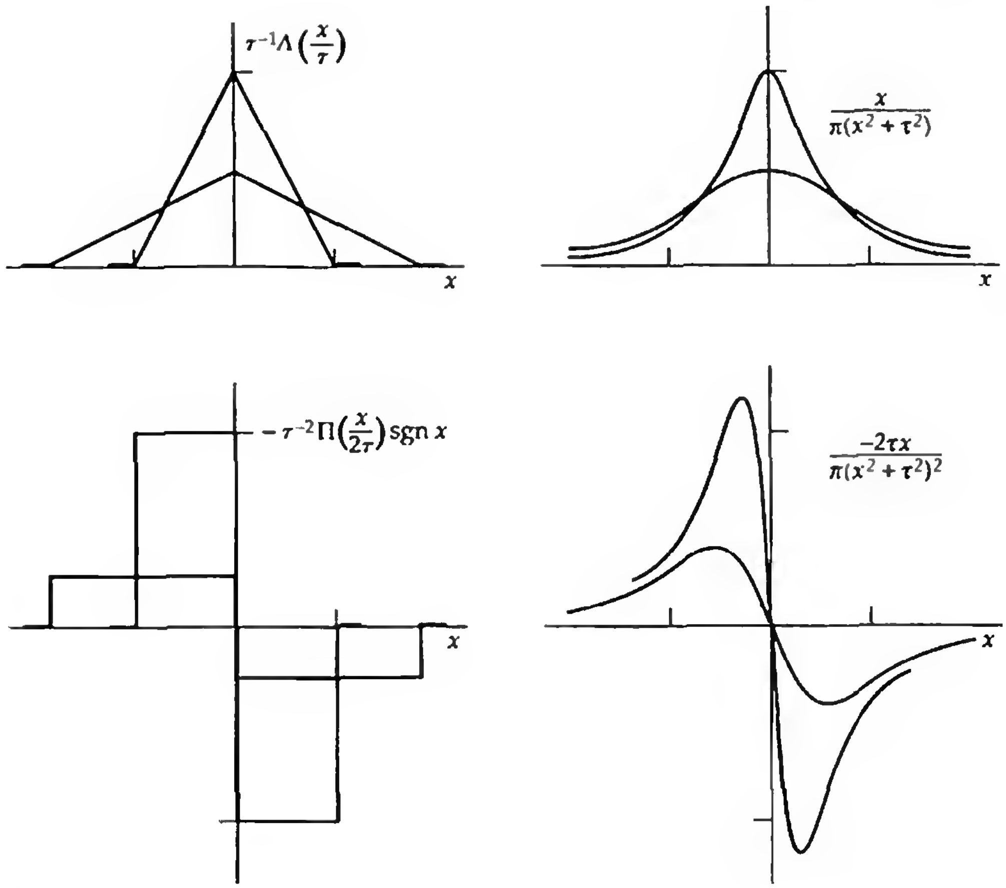


Fig. 5.9 **Pulse sequences (above) and their derivatives (below) which, as $\tau \to 0$, are used for contemplating the meaning of $\delta(x)$ and $\delta'(x)$.**

where the quantity in brackets is the derivative of one of the pulse shapes considered previously in discussing $\delta(x)$. The precise form of pulse adopted is unimportant—even a rectangular one will do—but in later work the possibility that a differentiable pulse shape may offer an advantage should be considered.

A derivative-sifting property

$$\delta' * f \equiv \int_{-\infty}^{\infty} \delta'(x - x')f(x')\,dx' = f'(x)$$

may be established in this way. Further properties are

$$\int_{-\infty}^{\infty} x\,\delta'(x)\,dx = -1$$
$$\int_{-\infty}^{\infty} |\delta'(x)|\,dx = \infty$$
$$x^2\,\delta'(x) = 0$$
$$\delta'(-x) = -\delta'(x) \qquad x\,\delta'(x) = -\delta(x)$$
$$f(x)\,\delta'(x) = f(0)\,\delta'(x) = f'(0)\,\delta(x).$$

The following relations apply to derivatives of higher order.

$$\int_{-\infty}^{\infty} \delta''(x)\,dx = 0$$

$$\delta''(x) * f(x) = f''(x)$$

$$\int_{-\infty}^{\infty} x^2\,\delta''(x)\,dx = 2$$

$$x^n\,\delta^{(n)}(x) = (-1)^n n!\,x^n\,\delta(x)$$

$$\delta^{(n)}(x) * f(x) = f^{(n)}(x)$$

$$\int_{-\infty}^{\infty} \delta^{(n)}(x) f(x)\,dx = (-1)^n f^{(n)}(0)$$

NULL FUNCTIONS

Null functions are known chiefly for having Fourier transforms which are zero, while not themselves being identically zero. By definition, $f(x)$ is a null function if

$$\int_a^b f(x)\,dx = 0$$

for all a and b. An alternative statement is

$$\int_{-\infty}^{\infty} |f(x)|\,dx = 0.$$

Null functions arise in connection with the one-to-one relationship between a function and its transform, a relationship defined by Lerch's theorem, which states that if two functions $f(x)$ and $g(x)$ have the same transform, then $f(x) - g(x)$ is a null function.

An example of a null function (see Fig. 5.10) is $\delta^0(x)$, an ordinary single-valued function defined by

$$\delta^0(x) = \begin{cases} 0 & x \neq 0 \\ 1 & x = 0, \end{cases}$$

which thus has a discontinuity at $x = 0$ similar in a way to that possessed by $H(x)$ [defined so that $H(0) = \frac{1}{2}$]. Under the ordinary rules of integration, the integral of $\delta^0(x)$ is certainly zero. However, it aptly describes the current taken by a series combination of a resistance and a capacitance from a battery, in the limit as the capacitance approaches zero.

We can now more succinctly state the relation between $H(x)$ and the step function $\hat{H}(x)$ of Fig. 4.14; thus

$$H(x) = \hat{H}(x) + \tfrac{1}{2}\delta^0(x),$$

the difference between the two being a null function.

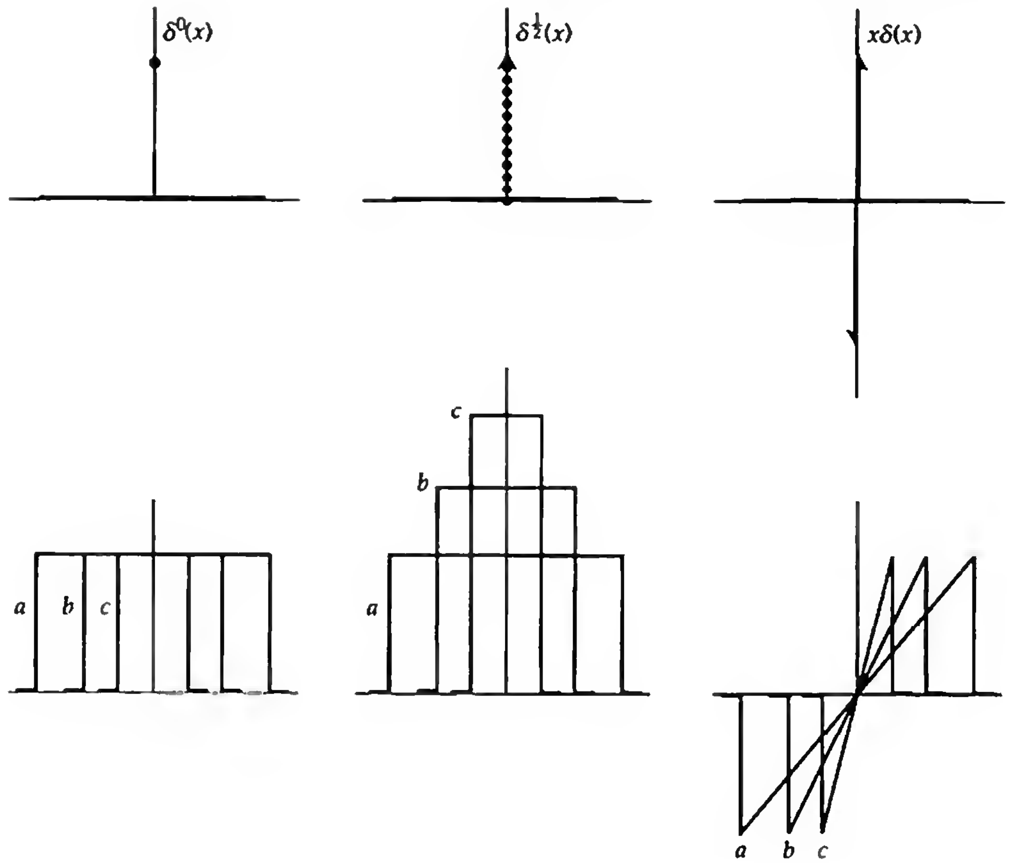


Fig. 5.10 Sequences (below) defining $\delta^0(x)$, $\delta^{\frac{1}{2}}(x)$, and $x\,\delta(x)$ (above).

The symbol $\delta^{\frac{1}{2}}(x)$ has to be considered in terms of sequences of pulses in the same way as $\delta(x)$. Consider the sequence

$$\tau^{-\frac{1}{2}}\Pi\left(\frac{x}{\tau}\right)$$

as $\tau \to 0$. Then we can attach meaning to the statements

$$\delta^{\frac{1}{2}}(x) = \begin{cases} 0 & x \neq 0 \\ \infty & x = 0, \end{cases}$$

$$\int_{-\infty}^{\infty} \delta^{\frac{1}{2}}(x)\,dx = 0,$$

and

$$\int_{-\infty}^{\infty} [\delta^{\frac{1}{2}}(x)]^2\,dx = 1.$$

We could describe $\delta^{\frac{1}{2}}(x)$ as a null symbol.

Exercise. Would we wish to call $\delta'(x)$ a null symbol?

SOME FUNCTIONS IN TWO AND MORE DIMENSIONS

One encounters the two- and three-dimensional impulse symbols ${}^2\delta(x,y)$, ${}^3\delta(x,y,z)$, as natural generalizations of $\delta(x)$. For example, ${}^2\delta(x,y)$ describes the pressure distribution over the (x,y)-plane when a concentrated unit force is applied at the origin; ${}^3\delta(x,y,z)$ describes the charge density in a volume containing a unit charge at the point (0,0,0). In establishing properties of ${}^2\delta(x,y)$ one considers a sequence, as $\tau \to 0$, of functions such as $\tau^{-2}\Pi(x/\tau)\Pi(y/\tau)$ or $(4/\pi)\tau^{-2}\Pi[(x^2+y^2)^{\frac{1}{2}}/\tau]$, which have unit volume (Fig. 5.11). Then we have

$$ {}^2\delta(x,y) = \begin{cases} 0 & x^2+y^2 \neq 0 \\ \infty & x^2+y^2 = 0,' \end{cases} $$

$$ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} {}^2\delta(x,y)\,dx\,dy = 1, $$

$$ {}^2\delta(ax,by) = \frac{1}{|ab|}\,{}^2\delta(x,y), $$

and the very interesting relation

$$ {}^2\delta(x,y) = \delta(x)\,\delta(y). $$

Introducing the radial coordinate r such that $r^2 = x^2 + y^2$, we can express ${}^2\delta(x,y)$ in terms of $\delta(r)$:

$$ {}^2\delta(x,y) = \frac{\delta(r)}{\pi|r|}. $$

In three dimensions,

$$ {}^3\delta(x,y,z) = \begin{cases} 0 & x^2+y^2+z^2 \neq 0 \\ \infty & x^2+y^2+z^2 = 0, \end{cases} $$

$$ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} {}^3\delta(x,y,z)\,dx\,dy\,dz = 1, $$

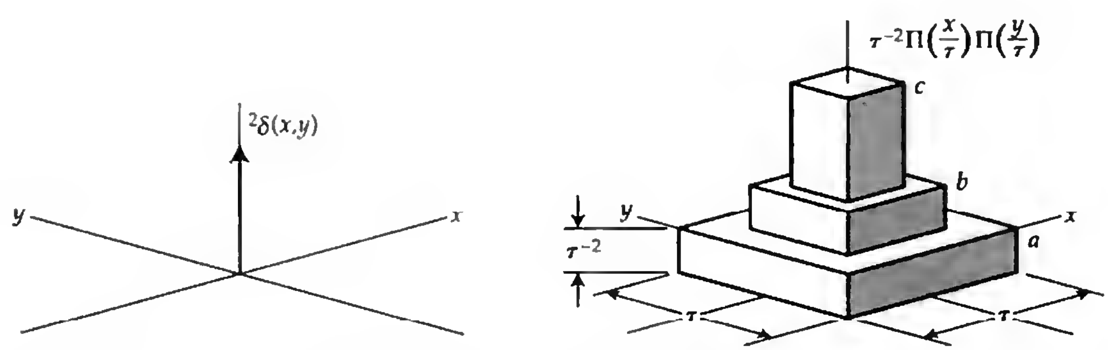


Fig. 5.11 **The two-dimensional impulse symbol ${}^2\delta(x,y)$ and a defining sequence of functions *a,b,c*.**

and $$^3\delta(x,y,z) = \delta(x)\,\delta(y)\,\delta(z) = {}^2\delta(x,y)\,\delta(z).$$

In cylindrical coordinates $r^2 = x^2 + y^2$,

$$^3\delta(x,y,z) = \frac{\delta(r)\,\delta(z)}{\pi|r|},$$

and with $\rho^2 = x^2 + y^2 + z^2$,

$$^3\delta(x,y,z) = \frac{\delta(\rho)}{2\pi\rho^2}.$$

For describing arrays in two dimensions we have the bed-of-nails symbol $^2\mathrm{III}(x,y)$, illustrated in Fig. 5.12 and defined by

$$^2\mathrm{III}(x,y) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} {}^2\delta(x-m,\,y-n).$$

Figure 5.13 shows an approach to the discussion of its properties. It has the property

$$^2\mathrm{III}(x,y) = \mathrm{III}(x)\,\mathrm{III}(y)$$

and various extensions of the integral properties of $^2\delta(x,y)$ and $\mathrm{III}(x)$, for example,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,{}^2\mathrm{III}(x,y)\,dx\,dy = \sum_m \sum_n f(m,n).$$

It is doubly periodic,

$$^2\mathrm{III}(x+m,\,y+n) = {}^2\mathrm{III}(x,y) \qquad m,n \text{ integral},$$

and $$\frac{1}{|XY|}\,{}^2\mathrm{III}\left(\frac{x}{X},\frac{y}{Y}\right)$$

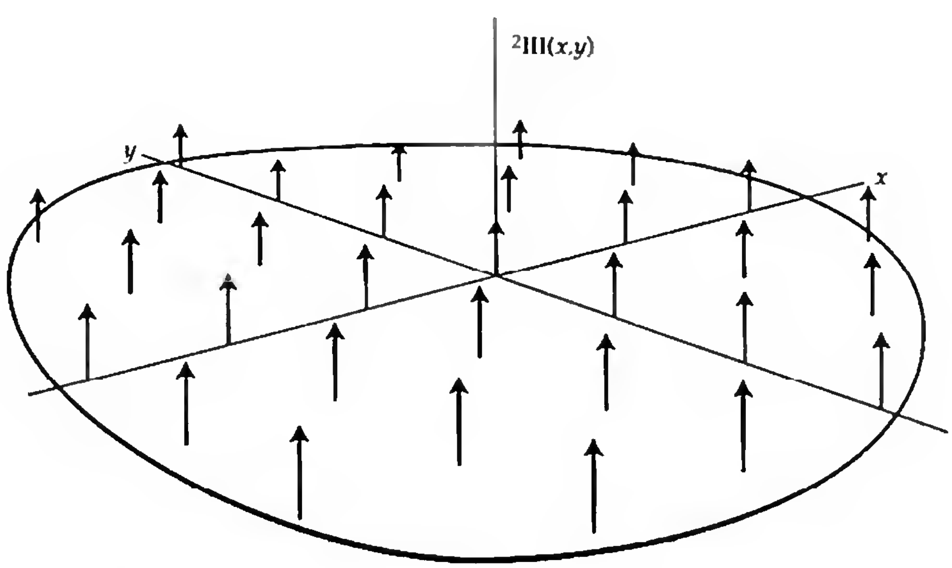


Fig. 5.12 The bed-of-nails symbol, $^2\mathrm{III}(x,y)$.

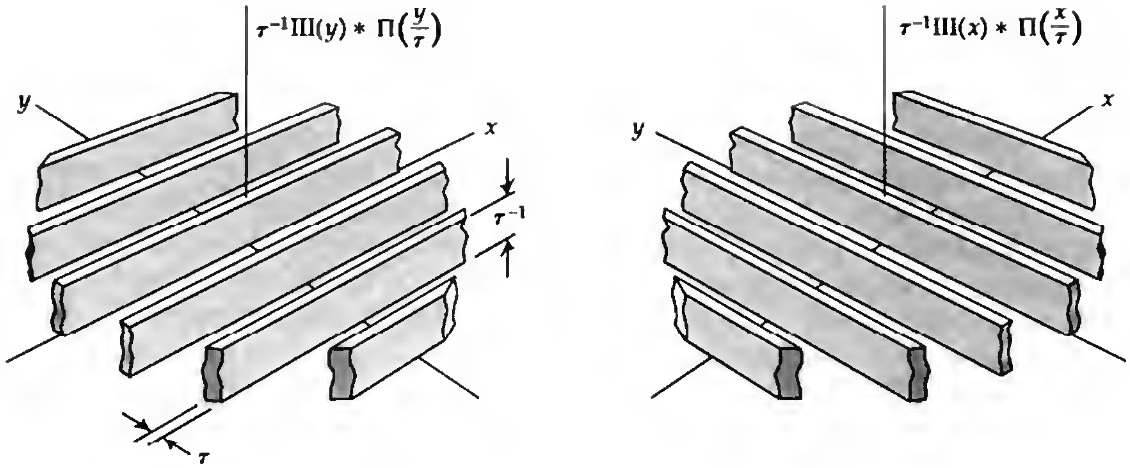


Fig. 5.13 Two functions whose product is suitable for discussing $^2\mathrm{III}(x,y)$.

represents a doubly periodic array of two-dimensional unit impulses with period X in the x direction and Y in the y direction.

Tabulation at discrete intervals of two independent variables (two-dimensionally sampled data), and the coefficients of double Fourier series are handled through the relation

$$f(x,y)\,^2\mathrm{III}(x,y) = \sum_m \sum_n f(m,n)\,^2\delta(x-m, y-n).$$

Convolution with $^2\mathrm{III}(x,y)$ describes replication in two dimensions, such as one has in a two-dimensional array of identical antennas, and, as with III in one dimension, $^2\mathrm{III}$ has the distinction of being its own two-dimensional Fourier transform (in the limit).

The scheme of multidimensional notation introduced here permits various self-explanatory extensions which are occasionally useful for compactness. Thus

$$^3\mathrm{III}(x,y,z) = \mathrm{III}(x)\mathrm{III}(y)\mathrm{III}(z)$$
$$^2\Pi(x,y) = \Pi(x)\Pi(y)$$
$$^2\mathrm{sinc}\,(x,y) = \frac{\sin \pi x \sin \pi y}{\pi^2 xy}$$
$$^2\mathrm{II}(x,y) = \mathrm{II}(x)\mathrm{II}(y).$$

Two other important two-dimensional distributions do not need new symbols. The row of spikes (see Fig. 5.14) is adequately expressed by $\mathrm{III}(x)\,\delta(y)$ and the grating by $\mathrm{III}(x)$. These two distributions form a two-dimensional Fourier transform pair and are suitable for discussing phenomena such as the diffraction of light by a row of pinholes or by a diffraction grating.

These two- and three-dimensional examples have further interpretations when derivatives of impulses are invoked, many of them familiar as line or sheet distributions of electric charge. For example, $\delta'(x)$ represents a line distribution

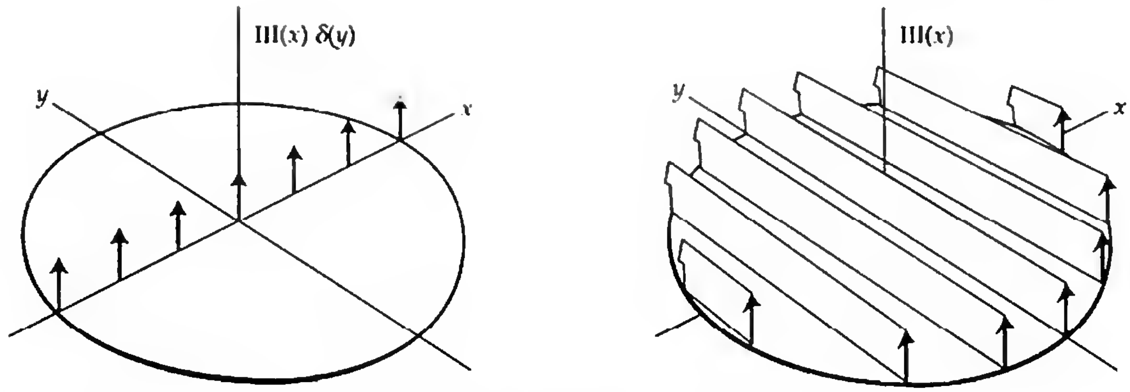


Fig. 5.14 **The row of spikes (left) and grating (right).**

of electric dipole moment confined to the y-axis on the (x, y)-plane. Interpreted in three-dimensional space, $\delta'(x)$ represents a sheet distribution of electric dipole moment on the (y, z)-plane. A simple dipole at (0, 0, 0) with unit moment in the x-direction is representable as $-\delta'(x)\delta(y)\delta(z)$. Quadrupole moment distributions are similarly expressible in terms of $\delta''(x)$. A simple unit quadrupole at (0, 0, 0), formed as the negative derivative with regard to x, is $\delta''(x)\delta(y)\delta(z)$. The negative derivative with respect to y produces a second kind of quadrupole. For example, $\delta'(x)\delta'(y)$ represents an electromagnetic radiator with a quatrefoil, or clover-leaf, radiation pattern in the (x, y)-plane. All these symbolic expressions may confidently be inserted into Maxwell's equations, the wave equation, Poisson's equation, and other fundamental differential equations.

THE CONCEPT OF GENERALIZED FUNCTION

As has been seen, a good deal of convenience attends the use of the impulse symbol $\delta(x)$ and other combinations of impulses such as $\text{III}(x)$ and $\text{II}(x)$. The word "symbol" has been used to call attention to the fact that these entities are not functions, but despite their apparent lack of status they have many uses, one of which is to provide derivatives for functions with simple discontinuities. Ordinarily we would say in such cases that the derivative does not exist, but the impulse symbol permits simple discontinuities to be accommodated.

The importance of dealing with discontinuous and impulsive behavior, even though it is nonphysical, was explained earlier in connection with the indispensable but nonphysical pure alternating and pure direct current.

Pure alternating current means an eternal harmonic variation, which cannot be generated. However, the response to a variation that is simple harmonic over a certain interval and zero outside that range can be made independent of the time of switching on, to a given precision, by waiting more than a certain length of time before observing the response. Since the details of time and manner of

switching on are irrelevant, they might as well be relegated to the infinitely remote past, thus making it convenient to refer to the observable response as the response to pure alternating current.

Impulses are likewise impossible to generate physically, but the responses to different sufficiently brief but finite pulses can be made indistinguishable to an instrument of given finite temporal resolving power.

Since precision of measurement is known to be limited by the temporal and spectral resolution of the measuring instrument, the physical limitations referred to in the preceding paragraphs can be characterized as "finite resolution." The feature of finite resolving power invoked in the explanation contains the key to the mathematical interpretation of impulse-symbol notation: integrals containing the impulse symbol are to be interpreted as limits of a sequence of integrals in which the impulse is replaced by a sequence of unit-area rectangular pulses $\tau^{-1}\Pi(x/\tau)$. The limit of the sequence of integrals may exist, even though the rectangular pulses grow *without* limit.

Thus the statement

$$\int_{-\infty}^{\infty} \delta(x) f(x)\, dx = f(0)$$

is deemed to mean

$$\lim_{\tau \to 0} \int_{-\infty}^{\infty} \tau^{-1}\Pi\left(\frac{x}{\tau}\right) f(x)\, dx = f(0).$$

In this interpretation of the statement the integrals can exist, and the limit of the integrals as $\tau \to 0$ can exist. The physical situation to which this corresponds is a sequence of ever more compact stimuli producing responses which become indistinguishable under observation to a given precision, no matter how high that precision is.

A satisfactory mathematical formulation of the theory of impulses has been evolved along these lines and is expounded in the books of Lighthill (1958) and Friedman (1956). Lighthill credits Temple with simplifying the mathematical presentation; Temple (1953) in turn credits the Polish mathematician Mikusiński (1948) with introducing the presentation in terms of sequences. Schwartz's two volumes (1950, 1951) on the theory of distributions unify "in one systematic theory a number of partial and special techniques proposed for the analytical interpretation of 'improper' or 'ideal' functions and symbolic methods" (Temple, 1953).

The idea of sequences was current in physical circles before 1948, however, going back to G.S. Kirchhoff in 1882 (van der Pol, 1937).

The introduction of rectangular pulse sequences was not meant to imply that other pulse shapes are not equally valid. In fact the essence of the approach is that the detailed pulse shape is unimportant. The advantage of rectangular pulses is the purely practical one of facilitating integration. However, rectangular pulses do not lend themselves to discussing the derivative of an impulse. For that we need something smoother which does not itself have an impulsive derivative.

Now for a general theory in which we wish to discuss derivatives of any order it is advantageous to have a pulse sequence such that derivatives of all orders exist. Schwartz and Temple introduce pulse shapes which have all derivatives and furthermore are zero outside a finite range; an example mentioned earlier in this chapter is

$$\begin{cases} \tau^{-1} e^{-\tau^2/(\tau^2 - x^2)} & |x| < \tau \\ 0 & |x| \geq \tau. \end{cases}$$

In actual fact one never inserts such a function explicitly into an integral; when it becomes necessary to integrate, a pulse shape with a *sufficient* number of derivatives is chosen. Often a rectangular pulse suffices.

Particularly well-behaved functions. The term "generalized function" may be defined as follows. First we consider the class S of functions which possess derivatives of all orders at all points and which, together with all the derivatives, die off at least as rapidly as $|x|^{-N}$ as $|x| \to \infty$, no matter how large N may be. We shall refer to members of the class S as particularly well-behaved functions. We note that the derivative and the Fourier transform of a particularly well-behaved function are also particularly well behaved. To prove the second of these statements let

$$\overline{F}(s) = \int_{-\infty}^{\infty} F(x) e^{-i2\pi sx}\, dx.$$

The conditions met by the particularly well-behaved function more than suffice to ensure that the Fourier integral exists. Differentiating p times, we have

$$\overline{F}^{(p)}(s) = \int_{-\infty}^{\infty} [(-i2\pi x)^p F(x)] e^{-i2\pi sx}\, dx,$$

and integrating by parts N times, we have

$$\begin{aligned} |\overline{F}^{(p)}(s)| &= \left| \frac{1}{(i2\pi s)^N} \int_{-\infty}^{\infty} \frac{d^N}{dx^N} [(-i2\pi x)^p F(x)] e^{-i2\pi sx}\, dx \right| \\ &\leq \frac{(2\pi)^{p-N}}{|s|^N} \int_{-\infty}^{\infty} \left| \frac{d^N}{dx^N} [x^p F(x)] \right| dx \\ &= O(|s|^{-N}); \end{aligned}$$

hence $\overline{F}(s)$ belongs to S.

Regular sequences. Among sequences of particularly well-behaved functions we distinguish sequences $p_\tau(x)$ which lead to limits when multiplied by any other particularly well-behaved function $F(x)$ and integrated. Thus if

$$\lim_{\tau \to 0} \int_{-\infty}^{\infty} p_\tau(x) F(x)\, dx$$

exists, then we call $p_\tau(x)$ a *regular* sequence of particularly well-behaved functions.

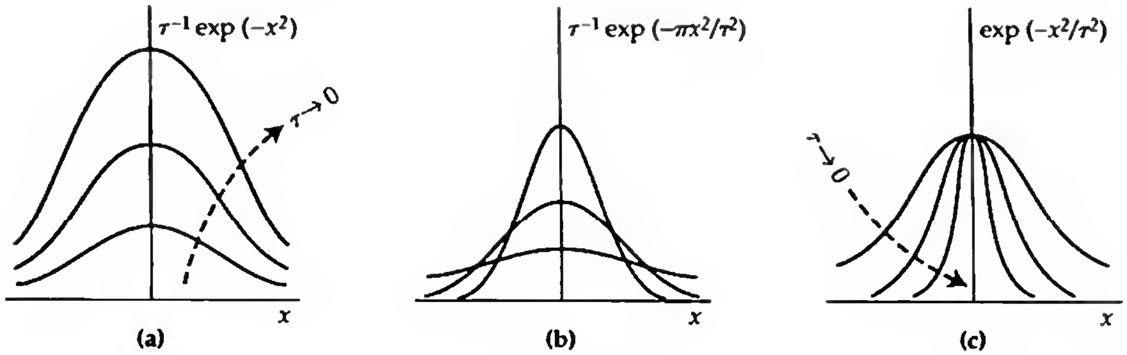


Fig. 5.15 **Sequences of particularly well-behaved functions: (a) not regular; (b) regular; (c) exercise.**

A sequence of particularly well-behaved functions which is not regular (see Fig. 5.15) is

$$\tau^{-1}e^{-x^2}.$$

An example of a regular sequence is

$$\tau^{-1}e^{-\pi x^2/\tau^2}.$$

In this example each member of the sequence has unit area, but that is not essential; for instance, consider the regular sequence $(1 + \tau^{-1})\exp(-x^2/\tau^2)$.

Generalized functions. A generalized function $p(x)$ is then taken to be defined by a regular sequence $p_\tau(x)$ of particularly well-behaved functions. In fact the generalized function *is* the regular sequence, and since the limit to which a regular sequence leads can be the same for more than one sequence, a generalized function is finally defined as the class of all regular sequences of particularly well-behaved functions equivalent to a given regular sequence. The symbol $p(x)$ thus represents an entity rather different from an ordinary function. It stands for a class of functions and is itself not a function. Therefore, when we write it in a context where ordinary functions are customary, the meaning to be assigned must be stated. For example, we shall deem that

$$\int_{-\infty}^{\infty} p(x)F(x)\,dx,$$

where $p(x)$ is a generalized function and $F(x)$ is any particularly well-behaved function, shall mean

$$\lim_{\tau\to 0}\int_{-\infty}^{\infty} p_\tau(x)F(x)\,dx,$$

where $p_\tau(x)$ is any regular sequence of particularly well-behaved functions defining $p(x)$.

Two points may be noticed. First, the sequences $\tau^{-1}\Pi(x/\tau)$ and $\tau^{-1}\Lambda(x/\tau)$ do not define a generalized function in the present sense, for the members of these

sequences do not possess derivatives of all orders at all points. Second, the function $F(x)$ on which the limiting process is tested has to be particularly well-behaved.

These highly restrictive conditions enable one to make a logical development at the cost of appearing to exclude the simple sequences and simple functions we ordinarily handle. We must bear in mind, however, that where differentiability is not in question we may fall back on rectangular pulses, and that where only first or second derivatives are required, Π^{*2} and Π^{*3} suffice. The requirement on asymptotic behavior is met by the rectangular pulse. On the other hand, the sequence τ^{-1} sinc (x/τ), which satisfies the requirement on differentiability, does not die away quickly enough to be a regular sequence in the strict sense; nevertheless, it is usable when it enters into a product with a function which is zero outside a finite interval.

The advantage of the Gaussian pulse as a special but simple case of a particularly well-behaved function has been exploited systematically by Lighthill (1958), whose line of development is followed.

The sequence $\exp(-\tau^2x^2)$ defines a generalized function $I(x)$, for

$$\lim_{\tau\to 0}\int_{-\infty}^{\infty} e^{-\tau^2x^2}F(x)\,dx = \int_{-\infty}^{\infty} F(x)\,dx,$$

which integral exists. Hence we can make the following statement about $I(x)$:

$$\int_{-\infty}^{\infty} I(x)F(x)\,dx = \int_{-\infty}^{\infty} F(x)\,dx,$$

where $F(x)$ is any particularly well-behaved function.

The sequence $\tau^{-1}\exp(-\pi x^2/\tau^2)$ defines a generalized function, for

$$\lim_{\tau\to 0}\int_{-\infty}^{\infty}\tau^{-1}e^{-\pi x^2/\tau^2}F(x)\,dx$$

exists and is equal to $F(0)$, where $F(x)$ is any particularly well-behaved function. To prove this note that

$$\begin{aligned}\left|\int_{-\infty}^{\infty}\tau^{-1}e^{-\pi x^2/\tau^2}F(x)\,dx - F(0)\right| &= \left|\int_{-\infty}^{\infty}\tau^{-1}e^{-\pi x^2/\tau^2}[F(x)-F(0)]\,dx\right| \\ &\leq \max|F'(x)|\int_{-\infty}^{\infty}\tau^{-1}e^{-\pi x^2/\tau^2}|x|\,dx \\ &= \frac{\tau}{\pi}\max|F'(x)|,\end{aligned}$$

which approaches zero as $\tau\to 0$. The generalized function defined by this and equivalent sequences is called $\delta(x)$, and we can state immediately that

$$\int_{-\infty}^{\infty}\delta(x)F(x)\,dx = F(0),$$

where $F(x)$ is any particularly well-behaved function.

Algebra of generalized functions. We have introduced one rule for handling the symbol standing for a generalized function, namely, a rule giving the meaning of

$$\int_{-\infty}^{\infty} p(x)F(x)\,dx.$$

Further rules are needed for handling the symbols for generalized functions where they appear in other algebraic situations.

Let $p(x)$ and $q(x)$ be two generalized functions, defined by the regular sequences $p_\tau(x)$ and $q_\tau(x)$, respectively. Now consider the sequence $p_\tau(x) + q_\tau(x)$. First, we note that it is a sequence of particularly well-behaved functions. Next, we see whether the sequence is a regular one, that is, whether

$$\lim_{\tau\to 0} \int_{-\infty}^{\infty} [p_\tau(x) + q_\tau(x)]F(x)\,dx$$

exists, where $F(x)$ belongs to S. The integral splits into two terms, each of which has a limit, since $p_\tau(x)$ and $q_\tau(x)$ are by definition regular sequences. The sum of the two limits is the limit whose existence thus establishes that $p_\tau(x) + q_\tau(x)$ is a regular sequence that consequently defines a generalized function. This generalized function we would wish to assign as the denotation of

$$p(x) + q(x);$$

it remains only to verify that the result is the same irrespective of the choice of the defining sequences $p_\tau(x)$ and $q_\tau(x)$, and indeed we see that the defining sequences $p_\tau(x) + q_\tau(x)$ are equivalent, since the sum of the two limits is independent of the choice of $p_\tau(x)$ and $q_\tau(x)$.

We now have a meaning for the addition of generalized functions.

Let $p(x)$ be a generalized function defined by a sequence $p_\tau(x)$. From the formula for integration by parts,

$$\int_{-\infty}^{\infty} p'_\tau(x)F(x)\,dx = -\int_{-\infty}^{\infty} p_\tau(x)F'(x)\,dx,$$

where $F(x)$ is any particularly well-behaved function and so therefore is $F'(x)$. Since $F'(x)$ is a particularly well-behaved function, and since $p_\tau(x)$ is by definition a regular sequence, it follows that

$$-\lim_{\tau\to 0} \int_{-\infty}^{\infty} p_\tau(x)F'(x)\,dx$$

exists, hence

$$\lim_{\tau\to 0} \int_{-\infty}^{\infty} p'_\tau(x)F(x)\,dx$$

exists. Thus $p'_\tau(x)$ is a regular sequence of particularly well-behaved functions, and all such sequences are equivalent. To the generalized function so defined we assign the notation

$$p'(x).$$

This gives us a meaning for the derivative of a generalized function. Here is an example of a statement which can be made about the derivative $p'(x)$ of a gener-

alized function $p(x)$:

$$\int_{-\infty}^{\infty} p'(x)F(x)\,dx = -\int_{-\infty}^{\infty} p(x)F'(x)\,dx.$$

Similarly, $$\int_{-\infty}^{\infty} p^{(n)}(x)F(x)\,dx = (-1)^n \int_{-\infty}^{\infty} p(x)F^{(n)}(x)\,dx.$$

Since by definition $F^{(n)}(x)$ exists, however large n may be, it follows that we have an interpretation for the nth derivative of a generalized function, for any n.

Differentiation of ordinary functions. Generalized functions possess derivatives of all orders, and if an ordinary function could be regarded as a generalized function, then there would be a satisfactory basis for formulas such as

$$\frac{d}{dx}[H(x)] = \delta(x).$$

If $f(x)$ is an ordinary function and we form a sequence $f_\tau(x)$ such that

$$\lim_{\tau\to 0} \int_{-\infty}^{\infty} f_\tau(x)F(x)\,dx = \int_{-\infty}^{\infty} f(x)F(x)\,dx,$$

where $F(x)$ is any particularly well-behaved function, then the sequence defines a *generalized* function, which we may denote by the same symbol $f(x)$. The symbol $f(x)$ then has two meanings. We shall limit attention to functions $f(x)$ which as $|x|\to\infty$ behave as $|x|^{-N}$ for some value of N.

A suitable sequence $f_\tau(x)$ is given by

$$[\tau^{-1}e^{-\pi x^2/\tau^2}] * [f(x)e^{-\tau^2x^2}].$$

With this enlargement of the notion of generalized functions we can embrace the unit step function $H(x)$ as a generalized function and assign meaning to its derivative $H'(x)$. Thus

$$\begin{aligned}\int_{-\infty}^{\infty} H'(x)F(x)\,dx &= -\int_{-\infty}^{\infty} H(x)F'(x)\,dx \\ &= -\int_{0}^{\infty} F'(x)\,dx \\ &= \int_{\infty}^{0} F'(x)\,dx \\ &= F(0),\end{aligned}$$

but $$\int_{-\infty}^{\infty} \delta(x)F(x)\,dx = F(0),$$

hence $$H'(x) = \delta(x).$$

The generalized function $\delta(x)$ is thus the derivative of the generalized function $H(x)$, and this is how we interpret formulas such as $H'(x) = \delta(x)$; we take the symbol for an ordinary function such as $H(x)$ to stand for the corresponding generalized function.

BIBLIOGRAPHY

Friedman B.: "Principles and Techniques of Applied Mathematics," John Wiley & Sons, New York, 1956.

Lighthill, M. J.: "An Introduction to Fourier Analysis and Generalised Functions," Cambridge University Press, Cambridge, England, 1958.

Mikusiński, J. G.: Sur la méthode de généralisation de Laurent Schwartz et sur la Convergence Faible, *Fundamenta Mathematicae*, vol. 35, p. 235, 1948.

Schwartz, L.: "Théorie des distributions," vols. 1 and 2, Herman & Cie, Paris, 1950 and 1951.

Temple, G.: Theories and Applications of Generalised Functions, *J. Lond. Math. Soc.*, vol. 28, p. 181, 1953.

Van der Pol, B.: Discontinuous Phenomena in Radio Communication, *J. Inst. Elec. Engrs.*, vol. 81, p. 381, 1937.

PROBLEMS

1. What is the even part of

$$\delta(x + 3) + \delta(x + 2) - \delta(x + 1) + \tfrac{1}{2}\delta(x) + \delta(x - 1) - \delta(x - 2) - \delta(x - 3)?$$

2. Attempting to clarify the meaning of $\delta(xy)$, a student gave the following explanation. "Where u is zero, $\delta(u)$ is infinite. Now xy is zero where $x = 0$ and where $y = 0$; therefore $\delta(xy)$ is infinite along the x and y axes. Hence $\delta(xy) = \delta(x) + \delta(y)$." Explain the fallacy in this argument, and show that

$$\delta(xy) = \frac{\delta(x) + \delta(y)}{(x^2 + y^2)^{\frac{1}{2}}}.$$

3. Show that

$$\mathrm{II}(x) = \delta(2x^2 - \tfrac{1}{2})$$

and that

$$\delta(x^2 - a^2) = \tfrac{1}{2}|a|^{-1}\{\delta(x - a) + \delta(x + a)\}. \triangleright$$

4. Show that

$$\int_{-\infty}^{\infty} e^{-i2\pi xs}\, ds = \delta(x)$$

and that

$$\int_{-\infty}^{\infty} \delta(x)e^{i2\pi sx}\, dx = 1.$$

5. Show that

$$\delta(ax + b) = \frac{1}{|a|}\,\delta\left(x + \frac{b}{a}\right), \qquad a \neq 0.$$

6. If $f(x) = 0$ has roots x_n, show that

$$\delta[f(x)] = \sum_n \frac{\delta(x - x_n)}{|f'(x_n)|}$$

wherever $f'(x_n)$ exists and is not zero. Consider the ideas suggested by $\delta(x^3)$ and $\delta(\operatorname{sgn} x)$.

7. Show that

$$\pi\delta(\sin \pi x) = \text{III}(x)$$

and

$$\delta(\sin x) = \pi^{-1}\,\text{III}\left(\frac{x}{\pi}\right). \rhd$$

8. Show that

$$\text{III}(x) + \text{III}(x - \tfrac{1}{2}) = 2\,\text{III}(2x) = \text{III}(x) * 4\text{II}(2x - \tfrac{1}{2}).$$

9. Show that

$$\text{III}(x)\Pi\left(\frac{x}{8}\right) = \text{III}(x)\Pi\left(\frac{x}{7}\right) + \frac{\text{II}(x/8)}{8}$$

and also that

$$\text{III}(x)\left(\frac{x}{6\frac{1}{2}}\right) = \text{III}(x)\Pi\left(\frac{x}{7\frac{1}{2}}\right). \rhd$$

10. Can the following equation be correct?

$$x\,\delta(x - y) = y\,\delta(x - y).$$

11. Show that $\Lambda(x) * \sum_{-\infty}^{\infty} a_n\,\delta(x - n)$ is the polygon through the points (n,a_n).

12. Prove that

$$\delta'(-x) = -\delta'(x)$$
$$x\,\delta'(x) = -\delta(x).$$

Show also that

$$f(x)\,\delta'(x) = f(0)\,\delta'(x) - f'(0)\,\delta(x),$$

for example, by differentiating $f(x)\,\delta(x)$. $\rhd$

13. In attempting to show that $\delta'(x) = -\delta(x)/x$ a student presented the following argument. "A suitable sequence, as τ approaches zero, for defining $\delta(x)$ is $\tau/\pi(x^2 + \tau^2)$. Therefore a suitable sequence for $\delta'(x)$ is the derivative

$$\frac{d}{dx}\frac{\tau}{\pi(x^2 + \tau^2)} = \frac{-2\tau x}{\pi(x^2 + \tau^2)^2}$$
$$= \frac{-2x}{x^2 + \tau^2}\,\frac{\tau}{\pi(x^2 + \tau^2)}.$$

The second factor is the sequence for $\delta(x)$, and the first factor goes to $-2/x$ in the limit as τ approaches zero. Therefore $\delta'(x) = -2\delta(x)/x$." Explain the fallacy in this argument.

14. Show that

$$x^n\delta^{(n)}(x) = (-1)^n n!\delta(x)$$

and hence that $$x^2\,\delta''(x) = 2\delta(x)$$

and $$x^3\,\delta''(x) = 0.$$

15. The function $[x]$ is here defined as the mean of the greatest integer less than x and the greatest integer less than or equal to x. Show that

$$[x]' = \mathrm{III}(x)$$

and also that $$\frac{d}{dx}\{[x]H(x)\} = \mathrm{III}(x)H(x) - \tfrac{1}{2}\delta(x).$$

(The common definition of $[x]$ as the greatest integer less than x is not fully suitable for the needs of this exercise; the two definitions differ by the null function which is equal to $\frac{1}{2}$ for integral values of x and is zero elsewhere.)

16. The sawtooth function $Sa(x)$ is defined by $Sa(x) = [x] - x + \frac{1}{2}$. Show that

$$Sa'(x) = \mathrm{III}(x) - 1$$

and that $$\frac{d}{dx}[Sa(x)H(x)] = [\mathrm{III}(x) - 1]H(x).$$

17. Show that $\operatorname{sgn}^2 x = 1 - \delta^0(x)$.

18. The Kronecker delta is defined by

$$\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j. \end{cases}$$

Show that it may be expressed as a null function of $i - j$ as follows:

$$\delta_{ij} = \delta^0(i - j).$$

19. We wish to consider the suitability of a sequence of asymmetrical profiles, such as $\tau^{-1}\{\Lambda(x/\tau) + \frac{1}{2}\Lambda[(x - \tau)/\tau]\}$, for representing the impulse symbol. Discuss the sifting property that leads to a result of the form

$$\delta_a * f = \mu\,\delta_+ * f + \nu\,\delta_- * f,$$

where δ_a is a symbol based on the asymmetrical sequence, δ_+ is based on the sequence $\tau^{-1}\Pi[(x - \frac{1}{2}\tau)/\tau]$, and δ_- is based on the sequence $\tau^{-1}\Pi[(x + \frac{1}{2}\tau)/\tau]$ (τ positive). ▷

20. Prove the relation ${}^2\delta(x,y) = \delta(r)/\pi|r|$.

21. Illustrate on an isometric projection the meaning you would assign to $\mathrm{III}[(x^2 + y^2)^{\frac{1}{2}}]$. How would you express something which on this diagram would have the appearance of equally spaced concentric rings of equal height?

22. The function $f_r(x)$ is formed from $f(x)$ by reversing it; that is, $f_r(x) = f(-x)$. Show that the operation of forming f_r from f can be expressed with the aid of the impulse symbol by

$$f \star \delta$$

and hence that

$$(f \star \delta) \star \delta = f.$$

23. Under what conditions could we say that $(f \star \delta) \star \delta = f \star (\delta \star \delta)$?

24. All the sequences $f(x,\tau)$ given on page 76 have the property that $f(0,\tau)$ increases without limit as $\tau \to 0$. Show that $\frac{1}{2}\tau^{-1}\Lambda[(x/\tau) - 1] + \frac{1}{2}\tau^{-1}\Lambda[(x/\tau) + 1]$ is an equivalent sequence which, however, possesses a limit of zero, as $\tau \to 0$, for all x. Show that $f(0,\tau)$, far from needing to approach ∞ as $\tau \to 0$, may indeed approach $-\infty$. ▷

25. Show that

$$f(x)\,\delta''(x) = f(0)\,\delta''(x) - 2f'(0)\,\delta'(x) + f''(0)\,\delta(x)$$

and that in general

$$f(x) + \delta^{(n)}(x) = f(0)\,\delta^{(n)}(x) - \binom{n}{1} f'(0)\,\delta^{(n-1)}(x) + \cdots$$

$$- \binom{n}{n-1} f^{(n-1)}(0)\,\delta'(x) + f^{(n)}(0)\,\delta(x). \ ▷$$

26. Impulses and sequences. The delta function possesses a "sifting property"

$$L = \lim_{\tau \to 0} \int_{\infty}^{\infty} f(x)h(x,\tau)\,dx = f(0),$$

which arises from physical situations where the function $h(x,\tau)$ can be thought of as a sequence of functions of x generated as a parameter τ ranges through a series of diminishing constant values.

(*a*) Give an example of a function $h(x,\tau)$ for which the sifting property does not hold.

(*b*) What conditions are sufficient for $h(x,\tau)$ to meet in order for the sifting property to be true?

(*c*) 1. Write down a particular case of $h(x,\tau)$ that meets your conditions.
2. Using your $h(x,\tau)$, take

$$f(x) = \begin{cases} -1 & x \ge 0 \\ 1 & x < 0. \end{cases}$$

Does the sifting property hold true for this example?

(*d*) A particular $h(x,\tau)$ has the property that, if attention is fixed on a given value of x, then $\lim_{\tau \to 0} h(x,\tau) = 0$, and the same applies for *all* values of x. Could such a function $h(x,\tau)$ satisfy the sifting property?

(*e*) Give an $h(x,\tau)$ such that L becomes $f(2)$.

(*f*) Give an $h(x,\tau)$ such that L becomes $f'(0)$.

(*g*) Give an $h(x,\tau)$ such that $L = \int_0^{\infty} f(x)\,dx$.

(h) Give an $h(x,\tau)$ such that $L = \sum_{n=0}^{\infty} f(n)$ (n integral).

27. Limits. A linear time-invariant system having an impulse response $I(t)$ is excited by an input voltage $V_1(t,\tau)$, where

$$V_1(t,\tau) = [1 + \tau^{-2}(1 - 2\tau)|x|]\Pi\left(\frac{x}{2\tau}\right).$$

and produces a response $V_2(t,\tau)$.

(*a*) Is it true that $\lim_{\tau\to 0} V_2(t,\tau) = I(t)$?

(*b*) Is it true that $\lim_{\tau\to 0} V_1(t,\tau) = 0$?

(*c*) Is it true that $\lim_{\tau\to 0} \int_{-\infty}^{\infty} V_1(t,\tau)\,dt = 0$?

(*d*) Is it true that $\int_{-\infty}^{\infty} \lim_{\tau\to 0} V_1(t,\tau)\,dt = 0$?

28. Sequence defining $\delta(x)$. Construct a sequence of particularly well-behaved functions $f(x,\tau)$ that defines $\delta(x)$ but has the property that $\lim_{\tau\to 0} f(x,\tau) = 0$ for all x.

29. Generalized functions. Consider the sequence of functions $\tau^{-1}\cos(\pi x^2/4\tau^2)$ generated as $\tau \to 0$. For *all* values of x, the function value diverges in an oscillatory manner without limit. Could such a sequence exhibit the sifting property of an impulse at $x = 0$; that is, could it be true that

$$\lim_{\tau\to 0} \int_{-\infty}^{\infty} \tau^{-1} \cos\left(\frac{\pi x^2}{4\tau^2}\right) F(x)\,dx = F(0)?$$

30. Asymmetrical impulse. Show that $\delta_+(x)$ introduced in Problem 5.19 differs from $\delta(x)$ by the derivative of a null function (p. 87)

$$\delta_+(x) = \delta(x) - \frac{1}{2}\frac{d}{dx}\delta^0(x).$$

31. Energy of voltage impulse. A voltage $V(t) = A\delta(t)$ is applied to a resistance R.

(*a*) How much charge is passed through the resistor?

(*b*) How much energy is dissipated in the resistor?

32. Delta notation. A voltage $V(t)$ is applied to a resistance R for a finite length of time during which 2 joules of energy are transferred to the resistor. If the experiment is repeated with a stronger voltage for a shorter time, too short a time, in fact, to be of interest, how could $V(t)$ be written in δ notation? ▷

33. Product of delta symbols. No definition is given to a product of impulses in the one-dimensional theory, but in two dimensions products arise naturally and are readily interpretable. Consider $\delta(x)\,\delta(y)$, each factor being regarded as a function of two variables and describing straight blades of unit height on the (x,y) plane (p. 335). (Unit height means unit line density or unit double integral per unit arc length.) Evaluate $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \delta(x)\,\delta(y)\,dx\,dy$ by (1) substituting $\tau^{-1}\Pi(x/\tau)$ for $\delta(x)$ and $\tau^{-1}\Pi(y/\tau)$ for $\delta(y)$ in accordance with the rule of p. 76, (2) performing the integration, and (3) proceeding to the limit as $\tau \to 0$. Also show that when two blades intersect at an angle θ, the double integral is increased by a factor $1/\sin\theta$. ▷

34. **Autocorrelation of ring impulse.** A circular ring impulse of total strength $2\pi a$, described by $\delta(r-a)$, arises in optics in dealing with annular slits and also comes up in other fields. Show that

$$\delta(r-a) \star\star\, \delta(r-a) = \left(\frac{r}{2a}\right)^{-1}\left[1-\left(\frac{r}{2a}\right)^2\right]^{-\frac{1}{2}} \Pi\left(\frac{r}{4a}\right).$$

Graph this function of r and explain the principal features. For example, why does the autocorrelation become infinite at $r=0$ and $r=a$; why are these singularities unequal; why is the value exactly 2 at $r/2a = 2^{-\frac{1}{2}}$? Investigate the autocorrelation of $\Pi(r/21) - \Pi(r/19)$. What are the values at $r=0$; $r=20$; what is the minimum value; and at what value of r does it occur? (In this problem $\star\star$ stands for two-dimensional autocorrelation.)

35. **Delta notation.** As an exercise in delta function notation, evaluate the following integrals.

$$\int_{-\infty}^{\infty} \delta(\sin x)\Pi(x-\tfrac{1}{2})\,dx, \quad \int_{-\infty}^{\infty} \delta(\cos x)\Pi\left(\frac{x}{4}\right) dx, \quad \int_{-\infty}^{\infty} \delta(\sin 2x)\Pi\left(\frac{x}{4}\right) dx. \;\triangleright$$

36. **Integral of impulse.** If $\int_{-\infty}^{\infty} \delta(x)\,dx$ is agreed to be unity, is there any objection to $\int_0^{\infty} \delta(x)\,dx = \frac{1}{2}$? ▷

37. **Two variables.** Use the method for interpreting expressions containing delta functions to arrive at the meaning of delta (x, y). ▷

38. **Checking analysis by computer.** It has been suggested that $\mathrm{III}(x)\operatorname{sgn} x$ has Fourier transform $-i\cot\pi s$. The innermost pair of impulses $-\delta(x+1)+\delta(x-1)$ has FT $-2i\sin 2\pi s$; then transforming pair by pair suggests that

$$\sum_{k=1}^{\infty} -2i\sin 2\pi ks = -i\cot\pi s.$$

To check for a sign error, a missing factor of 2, or more serious errors, compute both sides for a single value of s; for example, does $\sin 1^\circ + \sin 2^\circ + \sin 3^\circ \ldots$ add up to $\frac{1}{2}\cot\frac{1}{2}^\circ$? Graph the sum of N terms versus N and discuss your finding. ▷

6

The Basic Theorems

A small number of theorems play a basic role in thinking with Fourier transforms. Most of them are familiar in one form or another, but here we collect them as simple mathematical properties of the Fourier transformation. Most of their derivations are quite simple, and their applicability to impulsive functions can readily be verified by consideration of sequences of rectangular or other suitable pulses. As a matter of interest, proofs based on the algebra of generalized functions as given in Chapter 5 are gathered for illustration at the end of this chapter.

The emphasis in this chapter, however, is on illustrating the *meaning* of the theorems and gaining familiarity with them. For this purpose a stock-in-trade of particular transform pairs is first provided so that the meaning of each theorem may be shown as it is encountered.

A FEW TRANSFORMS FOR ILLUSTRATION

Six transform pairs for reference are listed below. They are all well known, and the integrals are evaluated in Chapter 7; we content ourselves at this point with asserting that the following integrals may be verified.

$$\int_{-\infty}^{\infty} e^{-\pi x^2} e^{-i2\pi xs}\, dx = e^{-\pi s^2} \quad \text{and} \quad \int_{-\infty}^{\infty} e^{-\pi s^2} e^{+i2\pi sx}\, ds = e^{-\pi x^2}$$

$$\int_{-\infty}^{\infty} \operatorname{sinc} x\, e^{-i2\pi xs}\, dx = \Pi(s) \quad \text{and} \quad \int_{-\infty}^{\infty} \Pi(s) e^{+i2\pi sx}\, ds = \operatorname{sinc} x$$

$$\int_{-\infty}^{\infty} \operatorname{sinc}^2 x\, e^{-i2\pi xs}\, dx = \Lambda(s) \quad \text{and} \quad \int_{-\infty}^{\infty} \Lambda(s) e^{+i2\pi sx}\, ds = \operatorname{sinc}^2 x$$

Thus the transform of the Gaussian function is the same Gaussian function, the transform of the sinc function is the unit rectangle function, and the transform of the sinc^2 function is the triangle function of unit height and area.

These formulas are illustrated as the first three transform pairs in Fig. 6.1. Note that the second row of the figure, which says that $\Pi(s)$ is the transform of sinc x, could be supplemented by a second figure, with left and right graphs interchanged, which would say that sinc s is the transform of $\Pi(x)$. A consequence of the reciprocal property of the Fourier transformation, this extra figure would appear redundant. However, the statement

$$\Pi(s) = \int_{-\infty}^{\infty} \operatorname{sinc} x \, e^{-i2\pi xs} \, dx$$

has quite a different character from

$$\operatorname{sinc} s = \int_{-\infty}^{\infty} \Pi(x) e^{-i2\pi xs} \, dx.$$

The first statement tells us that the integral of the product of certain rather ordinary functions is equal to unity for absolute values of the constant s less than $\frac{1}{2}$. Whether it is equal to say 0.3 or 0.35, the value of the integral is unchanged. However, if $|s|$ exceeds $\frac{1}{2}$, the situation changes abruptly, because the integral now comes to nothing and continues to do so, regardless of the precise value of s. Thus

$$\int_{-\infty}^{\infty} \frac{\sin \pi x}{\pi x} e^{-i2\pi xs} \, dx = \begin{cases} 1 & |s| < \frac{1}{2} \\ 0 & |s| > \frac{1}{2}. \end{cases}$$

This rather curious behavior is typical of many situations where the Fourier integral connects ordinary, continuous, and differentiable functions, on the one hand, with awkward, abrupt functions requiring piecewise definition, on the other. The second statement may be rewritten

$$\frac{\sin \pi s}{\pi s} = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{-i2\pi xs} \, dx.$$

Here an elementary definite integral of the exponential function is equal to an ordinary function of the parameter s. Thus the direct and inverse transforms express different things. Two separate and distinct physical meanings will later be seen to be associated with each transform pair.

Three further transforms required for illustrating the basic theorems are transform pairs in the limiting sense discussed earlier. Taking the result for the Gaussian function, and making a simple substitution of variables, we have[1]

$$\int_{-\infty}^{\infty} e^{-\pi(ax)^2} e^{-i2\pi sx} \, dx = |a|^{-1} e^{-\pi(s/a)^2}.$$

As $a \to 0$, the right-hand side represents a defining sequence for $\delta(s)$; the left-hand side is the Fourier transform of what in the limit is unity. It follows that

1 is the Fourier transform in the limit of $\delta(s)$.

[1]In this formula the absolute value of a is used in order to counteract the sign reversal associated with the interchange of the limits of integration when a is negative.

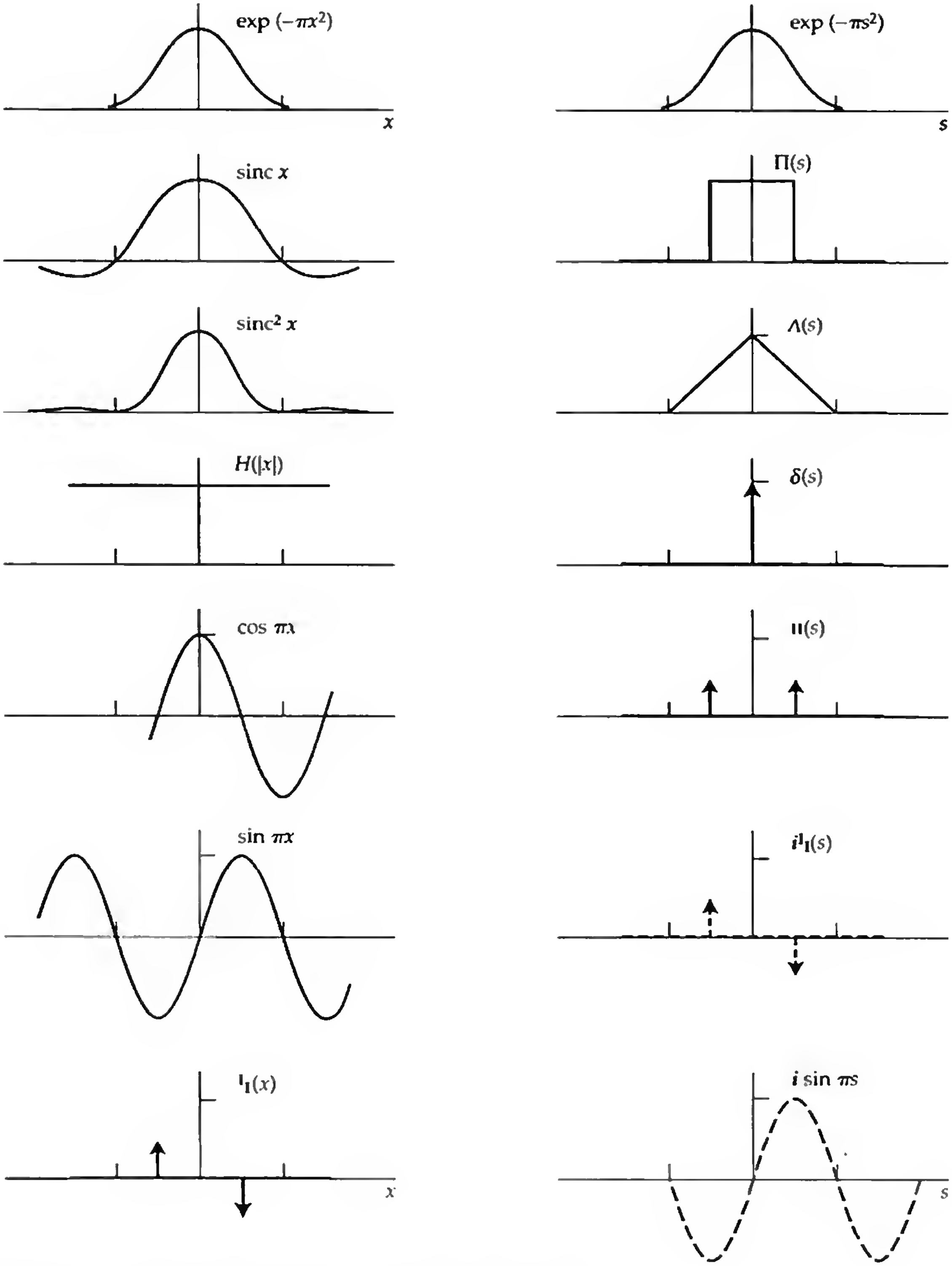


Fig. 6.1 Some Fourier transform pairs for reference.

The remaining two examples come from the verifiable relation

$$\int_{-\infty}^{\infty} e^{i\pi x} e^{-(ax)^2} e^{-i2\pi sx}\, dx = \frac{\sqrt{\pi}}{a}\, e^{-\pi^2[(s-\frac{1}{2})/a]^2},$$

whence

$$e^{i\pi x} \text{ is the Fourier transform in the limit of } \delta(s-\tfrac{1}{2});$$

or, splitting the left-hand side into real and imaginary parts and the right-hand side into even and odd parts, cos πx is the Fourier transform in the limit of

$$\tfrac{1}{2}\delta(s+\tfrac{1}{2}) + \tfrac{1}{2}\delta(s-\tfrac{1}{2}) = \mathrm{II}(s)$$

and i sin πx is the minus-i Fourier transform in the limit of

$$-\tfrac{1}{2}\delta(s+\tfrac{1}{2}) + \tfrac{1}{2}\delta(s-\tfrac{1}{2}) = -\,{}_{\mathrm{I}}{}^{\mathrm{I}}(s).$$

Summarizing the examples,

$$\begin{aligned}
e^{-\pi x^2} &\supset e^{-\pi s^2}\\
\operatorname{sinc} x &\supset \Pi(s)\\
\operatorname{sinc}^2 x &\supset \Lambda(s)\\
1 &\supset \delta(s)\\
\cos \pi x &\supset \mathrm{II}(s) \equiv \tfrac{1}{2}\delta(s+\tfrac{1}{2}) + \tfrac{1}{2}\delta(s-\tfrac{1}{2})\\
\sin \pi x &\supset i\,{}_{\mathrm{I}}{}^{\mathrm{I}}(s) \equiv \tfrac{1}{2}i\delta(s+\tfrac{1}{2}) - \tfrac{1}{2}i\delta(s-\tfrac{1}{2})\\
{}_{\mathrm{I}}{}^{\mathrm{I}}(x) &\supset i \sin \pi s.
\end{aligned}$$

All the transform pairs chosen for illustration has physical interpretations, which will be brought out later. Many properties appear among the transform pairs chosen for reference, including discontinuity, impulsiveness, limited extent, nonnegativeness, and oddness. The only examples exhibiting complex or nonsymmetrical properties are

$$e^{i\pi x} \supset \delta(s-\tfrac{1}{2})$$

and

$$\delta(x-\tfrac{1}{2}) \supset e^{-i\pi s}.$$

SIMILARITY THEOREM

If $f(x)$ has the Fourier transform $F(s)$, then $f(ax)$ has the Fourier transform $|a|^{-1}F(s/a)$.

Derivation:

$$\begin{aligned}
\int_{-\infty}^{\infty} f(ax)e^{-i2\pi xs}\, dx &= \frac{1}{|a|}\int_{-\infty}^{\infty} f(ax)e^{-i2\pi(ax)(s/a)}\, d(ax)\\
&= \frac{1}{|a|} F\left(\frac{s}{a}\right).
\end{aligned}$$

This theorem is well known in its application to waveforms and spectra, where compression of the time scale corresponds to expansion of the frequency scale. However, as one member of the transform pair expands horizontally, the other not only contracts horizontally but also grows vertically in such a way as to keep constant the area beneath it, as shown in Fig. 6.2.

A special case of interest arises with periodic functions and impulses. As Fig. 6.3 shows, expansion of a cosinusoid leads simply to shifts of the impulses con-

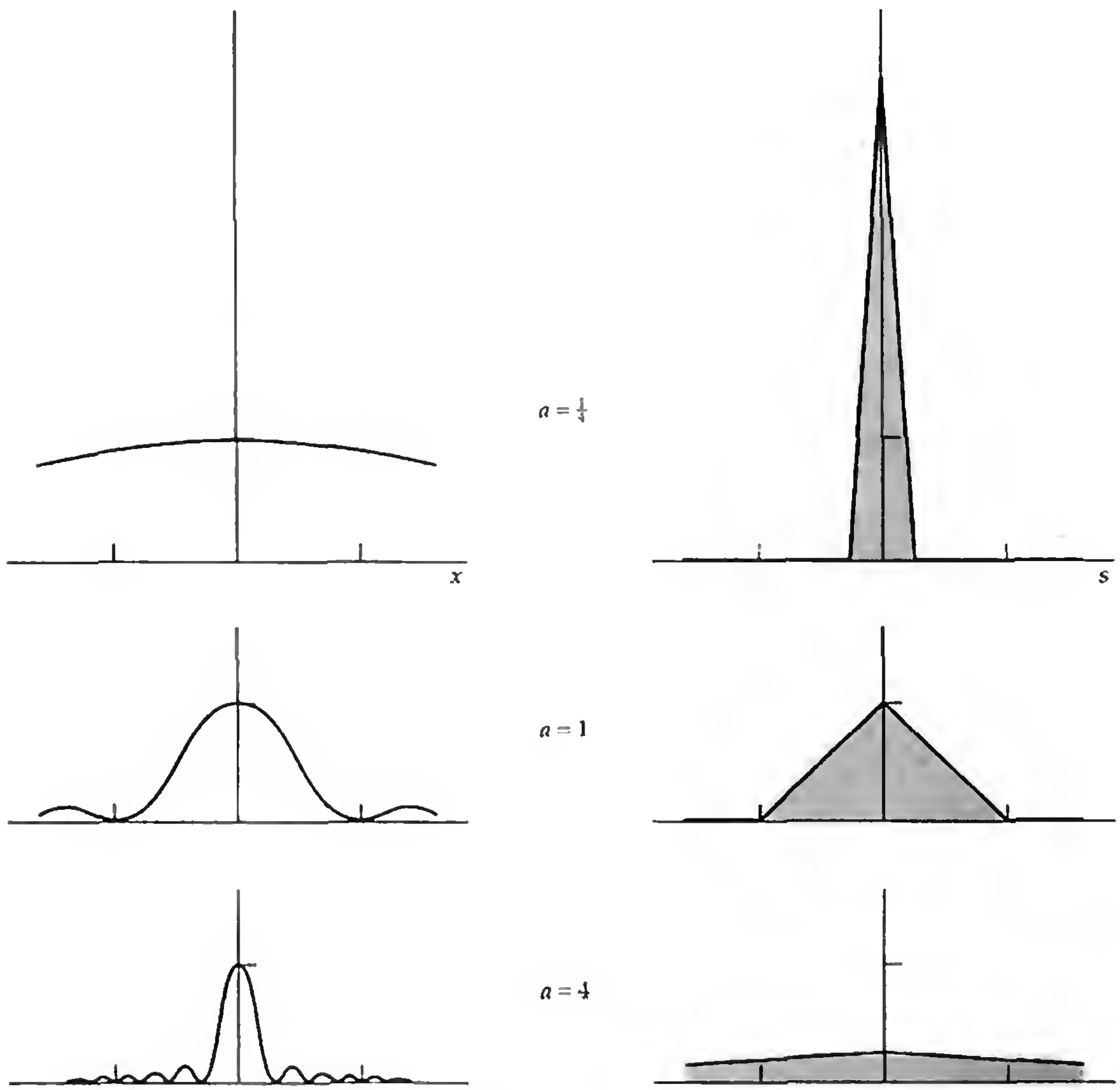


Fig. 6.2 The effect of changes in the scale of abscissas as described by the similarity, or abscissa-scaling, theorem. The shaded area remains constant.

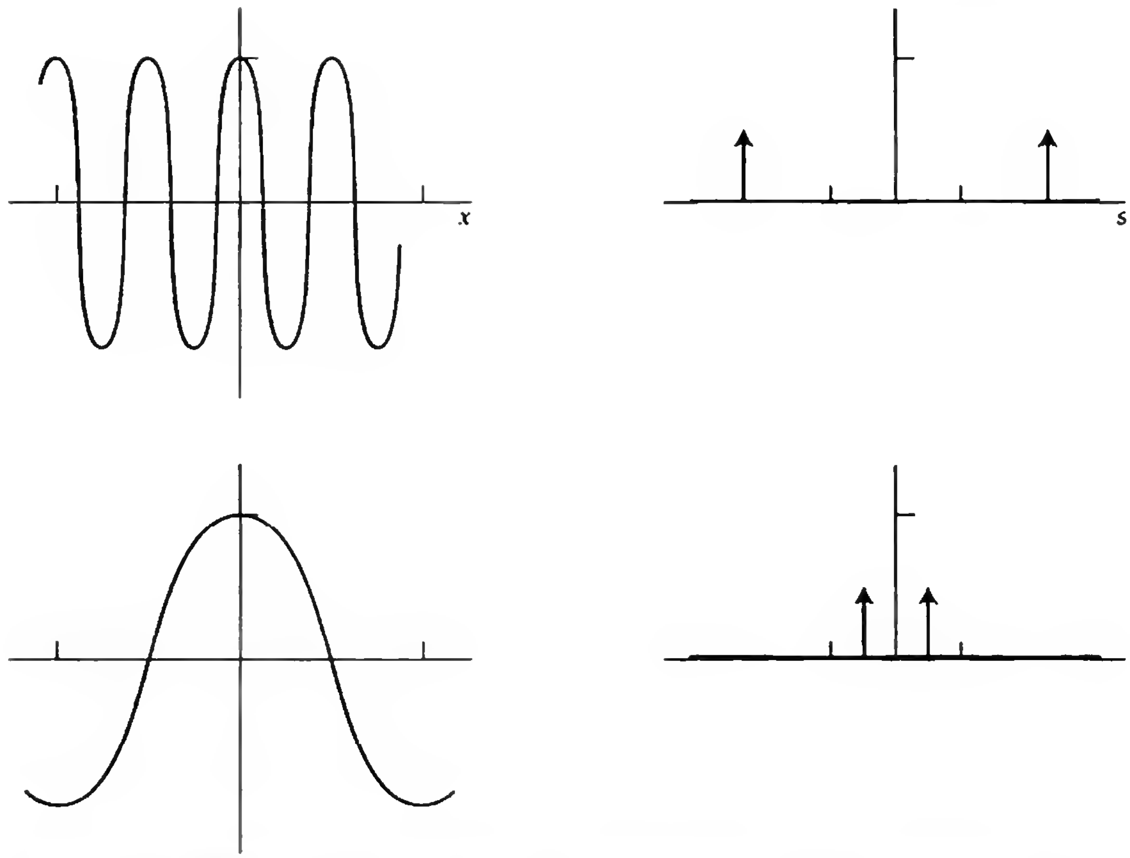


Fig. 6.3 **Expansion of a cosinusoid and corresponding shifts in its spectrum.**

stituting the transform. This is not simply a compression of the scale of s, for that would entail a reduction in strength of the impulses.

In a more symmetrical version of this theorem,

If $f(x)$ has the Fourier transform $F(s)$ then $|a|^{\frac{1}{2}}f(ax)$ has the Fourier transform $|b|^{\frac{1}{2}}F(bs)$, where $b = a^{-1}$.

Then, as each function expands or contracts it also shrinks or grows vertically (see Fig. 6.4) to compensate (in such a way that the integral of its square is maintained constant, as will be seen later from the power theorem).

ADDITION THEOREM

If $f(x)$ and $g(x)$ have the Fourier transforms $F(s)$ and $G(s)$, respectively, then $f(x) + g(x)$ has the Fourier transform $F(s) + G(s)$.

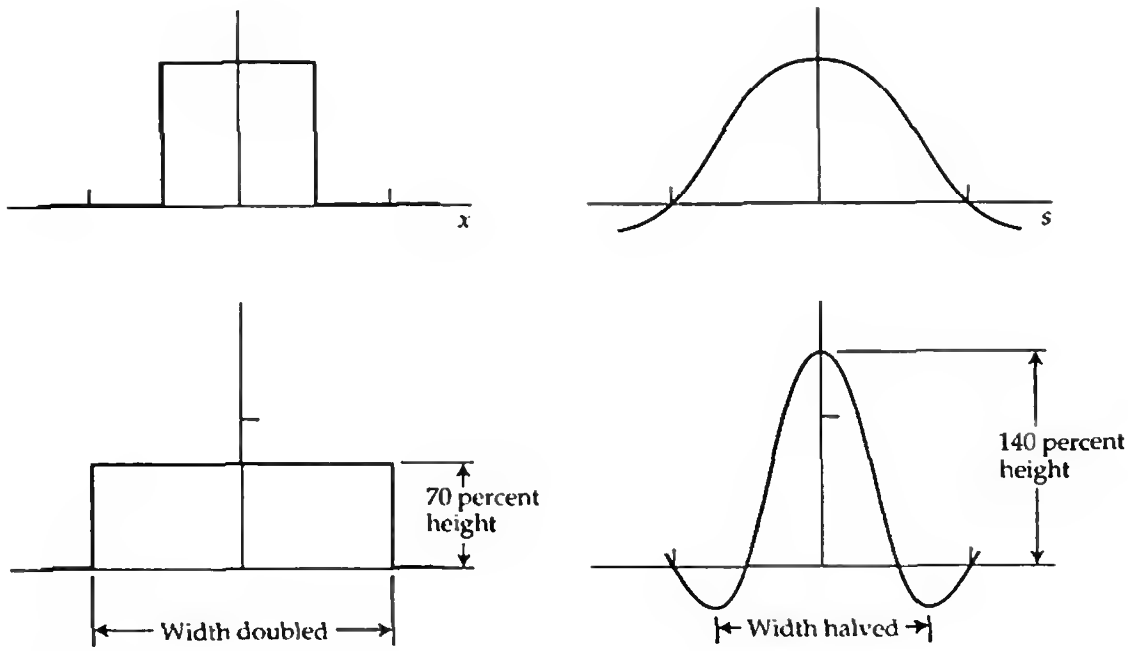


Fig. 6.4 **A symmetrical version of the similarity theorem.**

Derivation:

$$\int_{-\infty}^{\infty} [f(x) + g(x)]e^{-i2\pi xs}\,dx = \int_{-\infty}^{\infty} f(x)e^{-i2\pi xs}\,dx + \int_{-\infty}^{\infty} g(x)e^{-i2\pi xs}\,dx$$
$$= F(s) + G(s).$$

This theorem, which is illustrated by an example in Fig. 6.5, reflects the suitability of the Fourier transform for dealing with linear problems. A corollary is that $af(x)$ has the transform $aF(s)$, where a is a constant.

SHIFT THEOREM

If $f(x)$ has the Fourier transform $F(s)$, then $f(x - a)$ has the Fourier transform $e^{-2\pi ias}F(s)$.

Derivation:

$$\int_{-\infty}^{\infty} f(x - a)e^{-i2\pi xs}\,dx = \int_{-\infty}^{\infty} f(x - a)e^{-i2\pi(x-a)s}e^{-i2\pi as}d(x - a)$$
$$= e^{-i2\pi as}F(s).$$

If a given function is shifted in the positive direction by an amount a, no Fourier component changes in amplitude; it is therefore to be expected that the

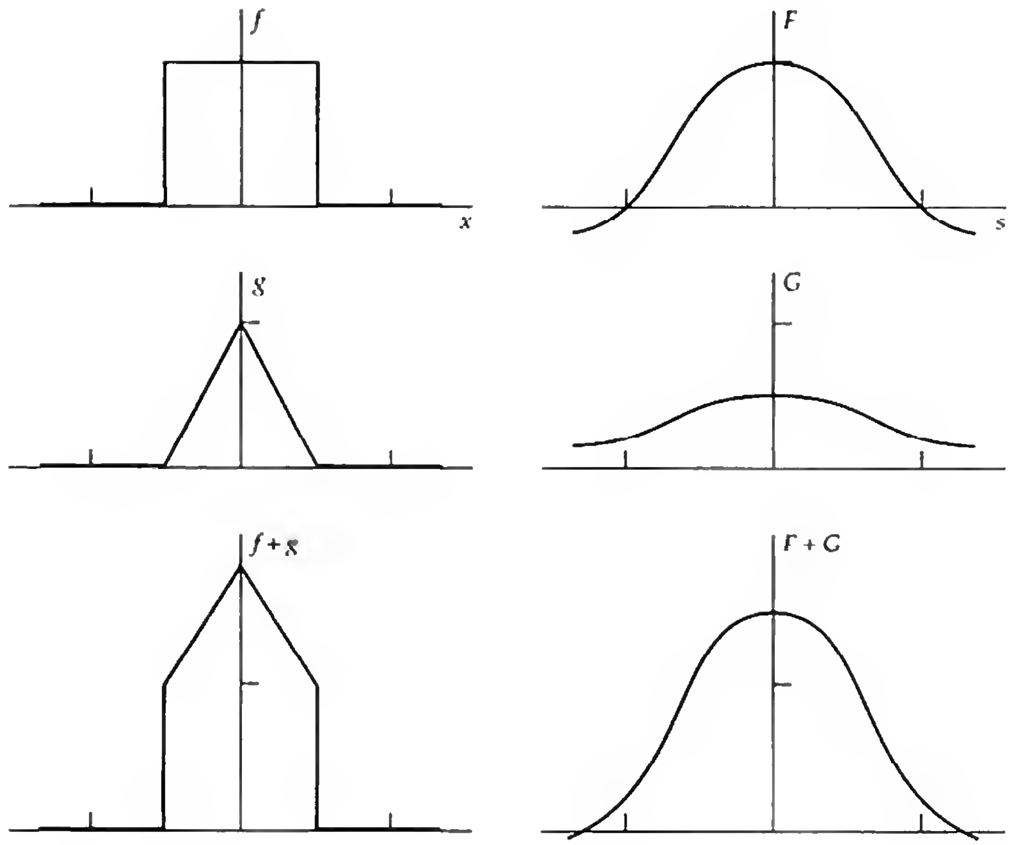


Fig. 6.5 **The addition theorem $f + g \supset F + G$.**

changes in its Fourier transform will be confined to phase changes. According to the theorem, each component is delayed in phase by an amount proportional to s; that is, the higher the frequency, the greater the change in phase angle. This occurs because the absolute shift a occupies a greater fraction of the period s^{-1} of a harmonic component in proportion to its frequency. Hence the phase delay is a/s^{-1} cycles or $2\pi as$ radians. The constant of proportionality describing the linear change of phase with s is $2\pi a$, the rate of change of phase with frequency being greater as the shift a is greater.

The shift theorem is one of those which are self-evident in a chosen physical embodiment. Consider parallel monochromatic light falling normally on an aperture. To shift the diffracted beam through a small angle, one changes the angle of incidence by that amount. But this is simply a way of causing the phase of the illumination to change linearly across the aperture; another way is to insert a thin prism that injects delays proportional to the prism thickness at each point of the aperture. These well-understood procedures for shifting the direction of a light beam are shown in Chapter 13 to exemplify the shift theorem.

In the example of Fig. 6.6, a function $f(x)$ is shown whose transform $F(s)$ is real. A shifted function $f(x - \frac{1}{4})$ has a transform which is derivable by subjecting $F(s)$ to a uniform twist of $\pi/2$ per unit of s. The figure attempts to show that the

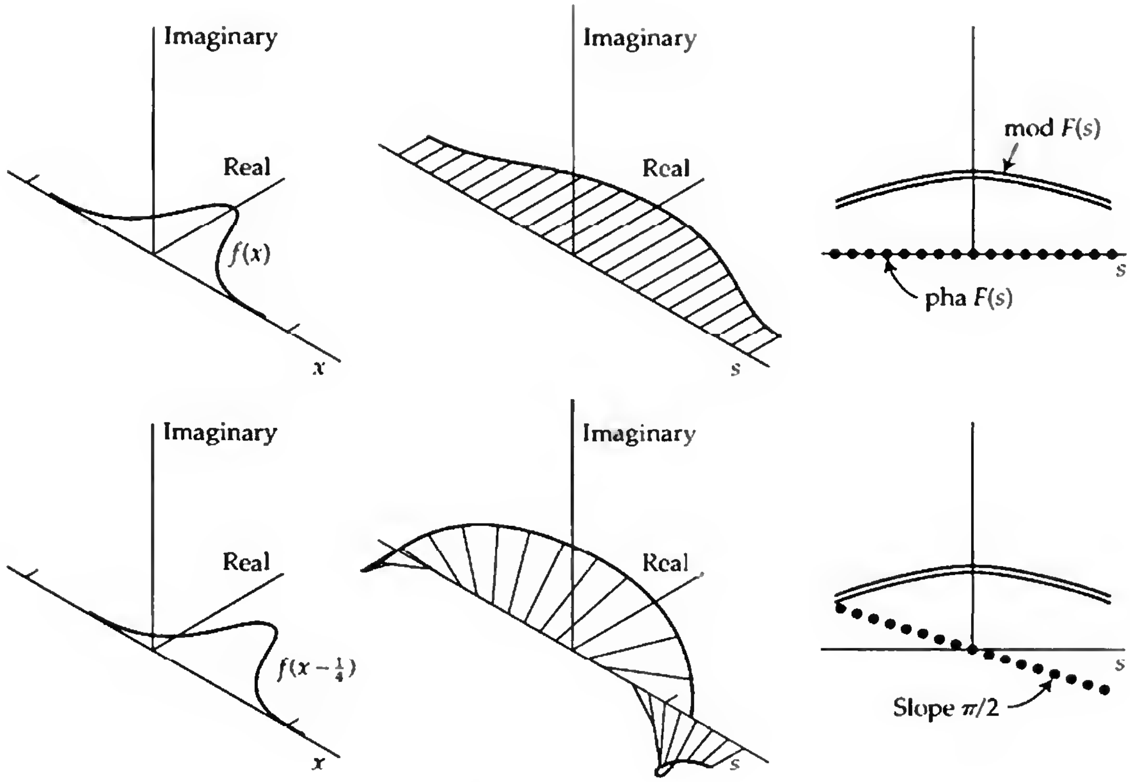


Fig. 6.6 **Shifting $f(x)$ by one quarter unit of x subjects $F(s)$ to a uniform twist of 90 deg per unit of s.**

plane containing $F(s)$ has been deformed into a helicoid. The practical difficulties of representing a complex function of s in a three-dimensional plot are overcome by showing the modulus and phase of $F(s)$ separately; however, the three-dimensional diagram often gives a better insight.

The second example (see Fig. 6.7) shows familiar results for the cosine and sine functions and for the intermediate cases which arise as the cosine slides along the axis of x. In this case the helicoidal surface is not shown. An alternative representation in terms of real and imaginary parts is given, incorporating the convention introduced earlier of showing the imaginary part by a broken line. A small shift evidently leaves the real part of the transform almost intact but introduces an odd imaginary part. With further shift the imaginary part increases until at a shift of $\pi/2$ there is no real part left. Then the real part reappears with opposite sign until at a shift of π both components have undergone a full reversal of phase.

MODULATION THEOREM

If $f(x)$ has the Fourier transform $F(s)$, then $f(x)\cos\omega x$ has the Fourier transform $\frac{1}{2}F(s-\omega/2\pi)+\frac{1}{2}F(s+\omega/2\pi)$.

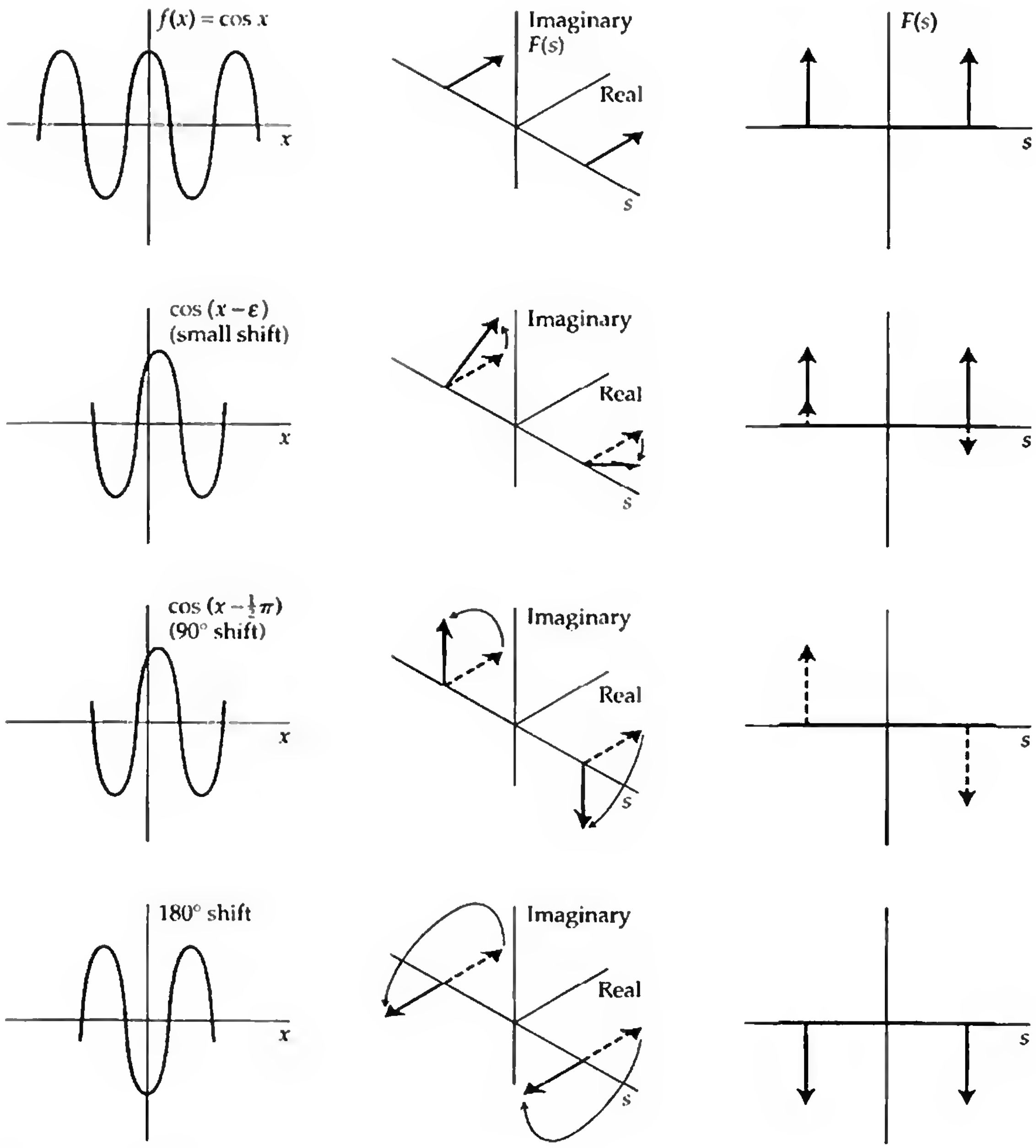


Fig. 6.7 The effect of shifting a cosinusoid.

Derivation:

$$\int_{-\infty}^{\infty} f(x) \cos \omega x \, e^{-i2\pi xs} \, dx = \frac{1}{2} \int_{-\infty}^{\infty} f(x) e^{i\omega x} e^{-i2\pi xs} \, dx + \frac{1}{2} \int_{-\infty}^{\infty} f(x) e^{-i\omega x} e^{-i2\pi xs} \, dx$$

$$= \frac{1}{2} \int_{-\infty}^{\infty} f(x) e^{-i2\pi(s - \omega/2\pi)x} \, dx + \frac{1}{2} \int_{-\infty}^{\infty} f(x) e^{-i2\pi(s + \omega/2\pi)x} \, dx$$

$$= \tfrac{1}{2} F(s - \omega/2\pi) + \tfrac{1}{2} F(s + \omega/2\pi).$$

The new transform will be recognized as the convolution of $F(s)$ with $\frac{1}{2}\delta(s + \omega/2\pi) + \frac{1}{2}\delta(s - \omega/2\pi) = (\pi/\omega)\text{II}(\pi s/\omega)$. This is a special case of the convolution theorem, but it is important enough to merit special mention. It is well known in radio and television, where a harmonic carrier wave is modulated by an envelope. The spectrum of the envelope is separated into two parts, each of half the original strength. These two replicas of the original are then shifted along the s axis by amounts $\pm\omega/2\pi$, as shown in Fig. 6.8.

CONVOLUTION THEOREM

As stated earlier, the convolution of two functions f and g is another function h defined by the integral

$$h(x) = \int_{-\infty}^{\infty} f(u)g(x - u)\, du.$$

A great deal is implied by this expression. For instance, $h(x)$ is a linear functional of $f(x)$; that is, $h(x_1)$ is a linear sum of values of $f(x)$, duly weighted as described by $g(x)$. However, it is not the most general linear functional; it is the particular kind for which any other value $h(x_2)$ is given by a linear combination of values of $f(x)$ weighted in the *same* way. Another way of conveying this special property of convolution is to say that a shift of $f(x)$ along the x axis results simply in an equal shift of $h(x)$; that is, if $h(x) = f(x) * g(x)$, then

$$f(x - a) * g(x) = h(x - a).$$

Suppose that a train is slowly crossing a bridge. The load at the point x is $f(x)$, and the deflection at x is $h(x)$. Since the structural members are not being pushed beyond the regime where stress is proportional to strain, it follows that the deflection at x_1 is a duly weighted linear combination of values of the load distribution $f(x)$. But as the train moves on, the deflection pattern does not move on with it unchanged; it is not expressible as a convolution integral. All that can be said in this case is that $h(x)$ is a linear functional of $f(x)$; that is,

$$h(x) = \int_{-\infty}^{\infty} f(u)g(x,u)\, du.$$

It is the property of *linearity* combined with *x-shift invariance* which makes Fourier analysis so useful; as shown in Chapter 9, this is the condition that simple harmonic inputs produce simple harmonic outputs with frequency unaltered.

If the well-known and widespread advantages of Fourier analysis are concomitant with the incidence of convolution, one may expect in the transform domain a simple counterpart of convolution in the function domain. This counterpart is expressed in the following theorem.

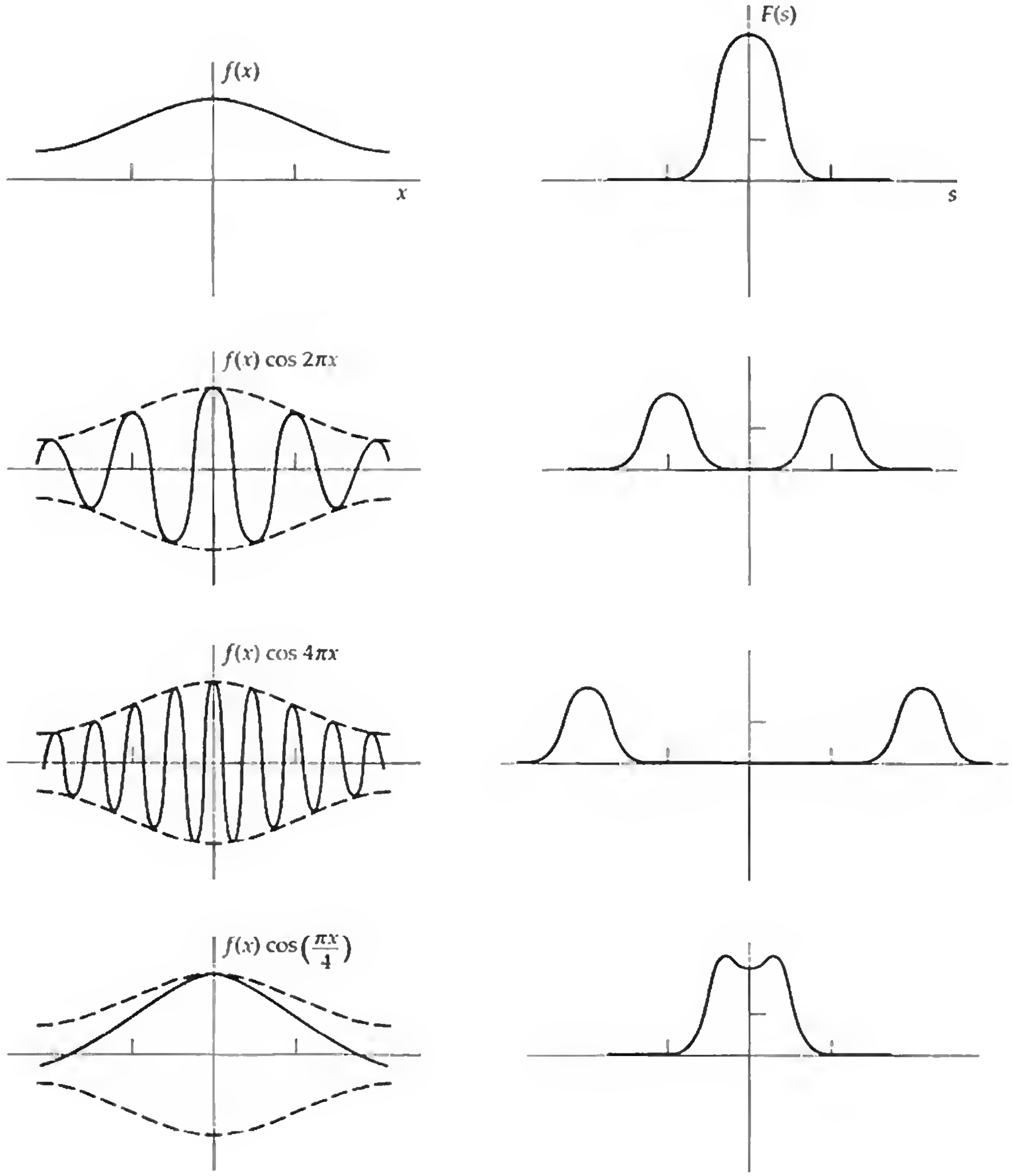


Fig. 6.8 An envelope function $f(x)$ multiplied by cosinusoids of various frequencies, with the corresponding spectra.

If $f(x)$ has the Fourier transform $F(s)$ and $g(x)$ has the Fourier transform $G(s)$, then $f(x) * g(x)$ has the Fourier transform $F(s)G(s)$; that is, convolution of two functions means multiplication of their transforms.

Derivation:

$$\begin{aligned}\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} f(x')g(x-x')\,dx'\right]e^{-i2\pi xs}\,dx \\ &= \int_{-\infty}^{\infty} f(x')\left[\int_{-\infty}^{\infty} g(x-x')e^{-i2\pi xs}\,dx\right]dx' \\ &= \int_{-\infty}^{\infty} f(x')e^{-i2\pi x's}G(s)\,dx' \\ &= F(s)G(s).\end{aligned}$$

Using bars to denote Fourier transforms, we can give compact statements of the theorem and its converse. Thus

$$\overline{f * g} = \bar{f}\bar{g},$$
$$\overline{fg} = \bar{f} * \bar{g}.$$

Equivalent statements are, using a long bar for the inverse transform,

$$\overline{\overline{f}\overline{g}} = f * g$$
$$\overline{\overline{f} * \overline{g}} = fg.$$

We have stated earlier that

$$\begin{aligned} f * g &= g * f && \text{(commutative)} \\ f * (g * h) &= (f * g) * h && \text{(associative)} \\ f * (g + h) &= f * g + f * h && \text{(distributive).} \end{aligned}$$

Further formulas are

$$\overline{f * g * h} = \bar{f}\bar{g}\bar{h},$$
$$\overline{f * (gh)} = \bar{f}(\bar{g} * \bar{h}).$$

This powerful theorem and its converse play an important role in transforming a function which can be recognized as the convolution of two others or as the product of two others.

The following are statements in words of some of the above equations.

1. The transform of a convolution is the product of the transforms.
2. The transform of a product is the convolution of the transforms.
3. The convolution of two functions is the transform of the product of their transforms.
4. The product of two functions is the transform of the convolution of their transforms.

Three valuable properties often used for checking are the following.

1. The area under a convolution is equal to the product of the areas under the "factors"; that is,

$$\int_{-\infty}^{\infty} (f * g)\, dx = \left[\int_{-\infty}^{\infty} f(x)\, dx\right]\left[\int_{-\infty}^{\infty} g(x)\, dx\right],$$

for
$$\int\left[\int f(u)g(x-u)\, du\right] dx = \int f(u)\left[\int g(x-u)\, dx\right] du$$
$$= \left[\int f(u)\, du\right]\left[\int g(x)\, dx\right].$$

2. The abscissas of the centers of gravity add; that is,

$$\langle x\rangle_{f*g} = \langle x\rangle_f + \langle x\rangle_g,$$

where
$$\langle x\rangle_h = \frac{\int_{-\infty}^{\infty} xh(x)\, dx}{\int_{-\infty}^{\infty} h(x)\, dx}.$$

3. The second moments add if $\langle x\rangle_f$ or $\langle x\rangle_g = 0$; in general,

$$\langle x^2\rangle_{f*g} = \langle x^2\rangle_f + \langle x^2\rangle_g + 2\langle x\rangle_f\langle x\rangle_g,$$

where
$$\langle x^2\rangle_h = \frac{\int_{-\infty}^{\infty} x^2h(x)\, dx}{\int_{-\infty}^{\infty} h(x)\, dx}.$$

It follows that the variances must add (p. 158).

We have enunciated the convolution theorem in the form

$f(x) * g(x)$ has the Fourier transform $F(s)G(s)$,

which, written in full, becomes either

$$\int_{-\infty}^{\infty} f(u)g(x-u)\, du = \int_{-\infty}^{\infty} F(s)G(s)e^{i2\pi xs}\, ds$$

or
$$F(s)G(s) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(u)g(x-u)e^{-i2\pi xs}\, du\, dx.$$

There are no fewer than 20 versions in which the convolution theorem is constantly needed, when we allow for complex conjugates and sign reversals of the variables. The 10 abbreviated forms are

$$\begin{aligned}
f * g &\supset FG \\
f * g(-) &\supset FG(-) \\
f(-) * g(-) &\supset F(-)G(-) \\
f * g^*(-) &\supset FG^* \\
f * g^* &\supset FG^*(-) \\
f(-) * g^*(-) &\supset F(-)G^* \\
f(-) * g^* &\supset F(-)G^*(-) \\
f^*(-) * g^*(-) &\supset F^*G^* \\
f^*(-) * g^* &\supset F^*G^*(-) \\
f^* * g^* &\supset F^*(-)G^*(-)
\end{aligned}$$

The self-convolution formulas are

$$\begin{aligned}
f * f &\supset F^2 \\
f(-) * f(-) &\supset [F(-)]^2 \\
f^*(-) * f^*(-) &\supset [F^*]^2 \\
f^* * f^* &\supset [F^*(-)]^2
\end{aligned}$$

and for autocorrelation we have

$$\begin{aligned}
f \star f &\supset FF(-) \\
f(-) \star f(-) &\supset FF(-) \\
f^*(-) \star f^*(-) &\supset F^*F^*(-) \\
f^* \star f^* &\supset F^*F^*(-)
\end{aligned}$$

RAYLEIGH'S THEOREM

The integral of the squared modulus of a function is equal to the integral of the squared modulus of its spectrum; that is,

$$\int_{-\infty}^{\infty} |f(x)|^2\,dx = \int_{-\infty}^{\infty} |F(s)|^2\,ds.$$

Derivation:

$$\begin{aligned}
\int_{-\infty}^{\infty} f(x)f^*(x)\,dx &= \int_{-\infty}^{\infty} f(x)f^*(x)e^{-i2\pi xs'}\,dx && s' = 0 \\
&= F(s') * F^*(-s') && s' = 0 \\
&= \int_{-\infty}^{\infty} F(s)F^*(s - s')\,ds && s' = 0 \\
&= \int_{-\infty}^{\infty} F(s)F^*(s)\,ds.
\end{aligned}$$

This theorem, which corresponds to Parseval's theorem for Fourier series, was first used by Rayleigh (1889) in his study of black-body radiation. In this, as in many other connections, each integral represents the amount of energy in a sys-

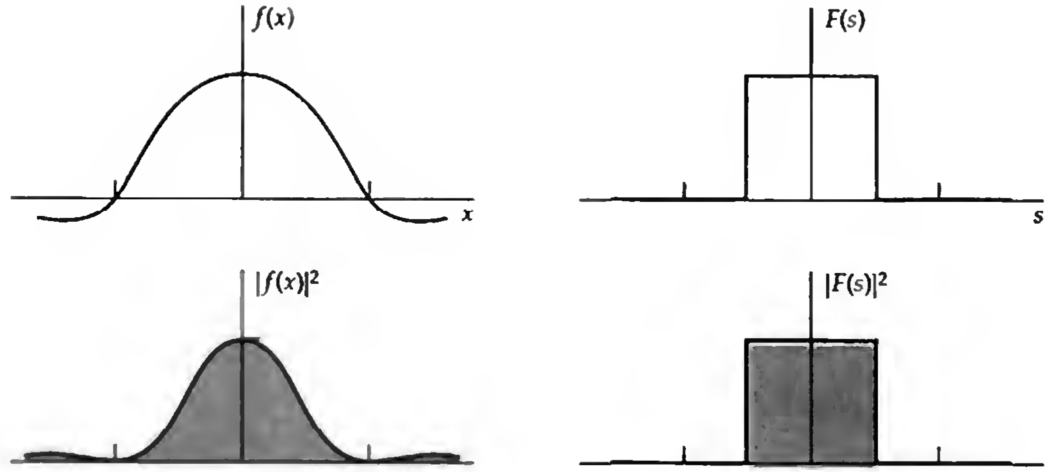


Fig. 6.9 **Rayleigh's theorem: the shaded areas are equal.**

tem, one integral being taken over all values of a coordinate, the other over all spectral components (see Fig. 6.9).

The theorem is sometimes referred to in mathematical circles as Plancherel's theorem (see Titchmarsh, 1924) after M. Plancherel, who in 1910 established conditions under which the theorem is true. The theorem is true if both the integrals exist. More recently, it has been shown by Carleman (see Bibliography, Chapter 2) that the theorem is true if one of the integrals exists. Rayleigh simply assumed in his derivation that the integrals existed.

POWER THEOREM

$$\int_{-\infty}^{\infty} f(x)g^*(x)\,dx = \int_{-\infty}^{\infty} F(s)G^*(s)\,ds.$$

Derivation: The proof is as for Rayleigh's theorem when f^* is replaced by g^* and F^* by G^*. The following version illustrates a compact notation which is useful in its place.

$$\int fg^*\,dx = \overline{fg^*}|_0 = \overline{f} * \overline{g^*}|_0 = F * G^*(-)|_0 = \int FG^*\,ds.$$

In many physical interpretations, each side of this equation represents energy or power (see Fig. 6.10), two different approaches being used to evaluate the energy or power. In one approach the instantaneous or local power or energy is evaluated as the product of a pair of canonically conjugate variables (electric and magnetic fields, voltage and current, force and velocity) integrated over time or space. In the second approach the temporal or spatial spectral components are multiplied and integrated over the whole spectrum.

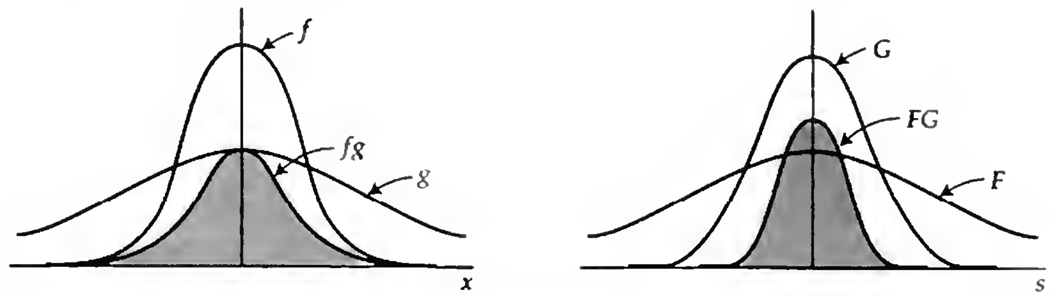


Fig. 6.10 **The power theorem: the shaded areas are equal. In this example f and g are real and even.**

It very often happens that both f and g are real quantities, as in the three examples cited. Then F and G may be complex, and

$$\begin{aligned} FG^* &= (\operatorname{Re} F + i \operatorname{Im} F)(\operatorname{Re} G - i \operatorname{Im} G) \\ &= (\operatorname{Re} F)(\operatorname{Re} G) + (\operatorname{Im} F)(\operatorname{Im} G) + \text{odd terms.} \end{aligned}$$

Inspection shows that the final terms are odd, for F and G are hermitian: that is, their real parts are even and imaginary parts odd. The odd terms do not contribute to the infinite integral. Hence for real f and g

$$\int_{-\infty}^{\infty} fg\, dx = \int_{-\infty}^{\infty} FG^*\, ds = \int_{-\infty}^{\infty} [(\operatorname{Re} F)(\operatorname{Re} G) + (\operatorname{Im} F)(\operatorname{Im} G)]\, ds.$$

This situation is illustrated in Fig. 6.11.

Exercise. Show that, provided f and g are real, an alternative version of the power theorem is

$$\int_{-\infty}^{\infty} f(x)g(-x)\, dx = \int_{-\infty}^{\infty} F(s)G(s)\, ds.$$

By putting $g(x) = f(x)$ we obtain Rayleigh's theorem, which is thus appropriate to physical systems where, f/g and F/G (often interpretable as impedance

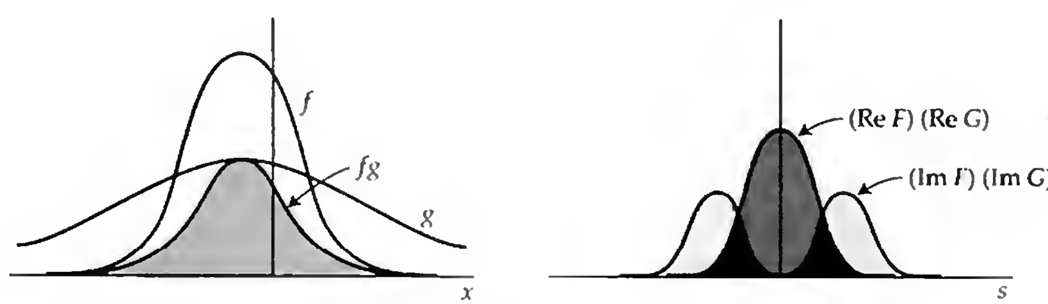


Fig. 6.11 **The power theorem for f and g real, F and G complex. The shaded area on the left equals the sum of the shaded areas on the right.**

in its general sense) being constant, energy or power may be expressed as the square of one variable alone. The theorem does not have a distinctive name of its own; some authors refer to it as Parseval's theorem, which is the well-established name of a theorem in the theory of Fourier series (Chapter 10).

AUTOCORRELATION THEOREM

If $f(x)$ has the Fourier transform $F(s)$, then its autocorrelation function $\int_{-\infty}^{\infty} f^*(u)\, f(u+x)\, du$ has the Fourier transform $|F(s)|^2$.

Derivation:

$$\begin{aligned}\int_{-\infty}^{\infty} |F(s)|^2 e^{i2\pi xs}\, ds &= \int_{-\infty}^{\infty} F(s)F^*(s)e^{i2\pi xs}\, ds \\ &= f(x) * f^*(-x) \\ &= \int_{-\infty}^{\infty} f(u)f^*(u-x)\, du \\ &= \int_{-\infty}^{\infty} f^*(u)f(u+x)\, du.\end{aligned}$$

A special case of the convolution theorem, the autocorrelation theorem is familiar in communications in the form that the autocorrelation function of a signal is the Fourier transform of its power spectrum.[2] It is illustrated in Fig. 6.12. The unique feature of this theorem, as contrasted with a theorem that could be stated for the self-convolution, is that information about the phase of $F(s)$ is entirely missing from $|F(s)|^2$. The autocorrelation function correspondingly contains no information about the phase of the Fourier components of $f(x)$, being unchanged if phases are allowed to alter, as was shown on p. 45.

Exercise. Show that the normalized autocorrelation function $\gamma(x)$, for which $\gamma(0) = 1$ (see p. 41), has as its Fourier transform the normalized power spectrum $|\Phi(s)|^2$ whose infinite integral is unity, and which is defined by

$$|\Phi(s)|^2 = \frac{|F(s)|^2}{\int_{-\infty}^{\infty} |F(s)|^2\, ds}.$$

A statement may also be added about the function $C(x)$, which was defined in Chapter 3 by the sequence of autocorrelation functions $\gamma_X(x)$ generated from

[2] The corresponding theorem for signals that do not tend to zero as time advances is sometimes referred to as Wiener's theorem (1949).

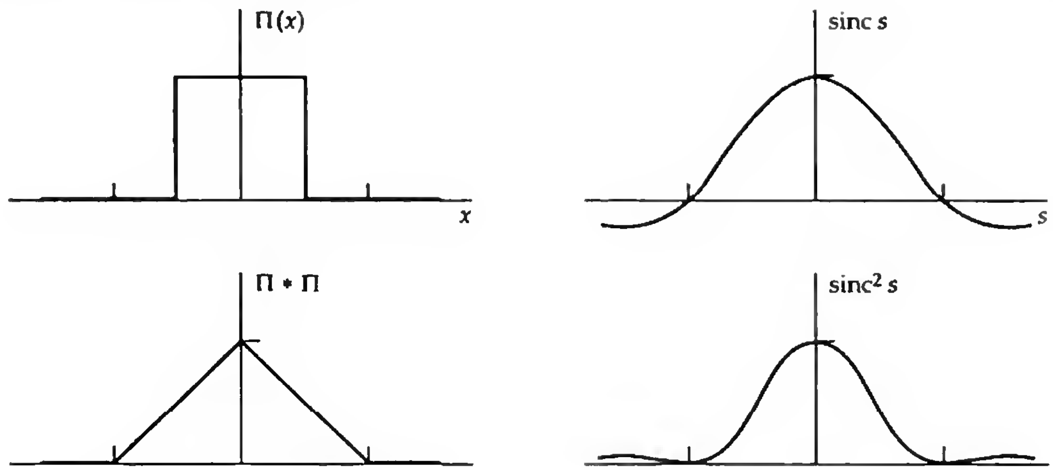


Fig. 6.12 **The autocorrelation theorem: autocorrelating a function corresponds to squaring (the modulus of) its transform.**

the functions $f(x)\Pi(x/X)$ as $X \to \infty$. If $\gamma_X(x)$ approached a limit, then the limit was called $C(x)$. Corresponding to the sequence of normalized autocorrelation functions, $\gamma_X(x)$ is the sequence of normalized power spectra $|\Phi_X(s)|^2$. If $\gamma_X(x)$ approaches a limit as the segment length X increases, then the normalized power spectrum settles down to a limiting form $|\Phi_\infty(s)|^2$. In these circumstances the autocorrelation theorem takes the form

$$C(x) \supset |\Phi_\infty(s)|^2$$

and one generally says, as before, that the autocorrelation is the Fourier transform of the power spectrum, suiting the definitions to the needs of the case.

Clearly it may happen that the sequence of transforms of $\gamma_X(x)$ does not approach limits for all s but is of a character describable with impulse symbols $\delta(s)$. Therefore situations may be entertained where the transform of $C(x)$ is a generalized function. For example, an ideal line spectrum such as is possessed by a signal carrying finite power at a single frequency is such a case. We know that if $f(x) = \cos \alpha x$, then $C(x) = \cos \alpha x$. The Fourier transform of $C(x)$ is thus a generalized function $\frac{1}{2}\delta(s + \alpha/2\pi) + \frac{1}{2}(s - \alpha/2\pi)$, and if necessary we could work out the sequence of transforms of $\gamma_X(x)$ that define it. The interesting point here, however, is that the power spectrum as a generalized function is not deducible from the autocorrelation theorem, for no interpretation has been given for products such as $[\delta(x)]^2$.

Exercise. Give an interpretation for $[\delta(x)]^2$ by attempting to apply the autocorrelation theorem to $f(x) = \cos \alpha x$ and test it on some other simple example such as $f(x) = 1$.

Exercise. Show that the situation cannot arise where the sequence $\gamma_X(x)$ calls for the use of $\delta(x)$ in representing $C(x)$.

Exercise. We wish to discuss the ideal situation of a power spectrum which is flat and extends to infinite frequency. Determine $C(x)$ and its transform. Show that the nonnormalized autocorrelation function lends itself to this requirement, and that a good version of the autocorrelation theorem can be devised in which the power spectrum is normalized so as to be equal to unity at its origin. What does this form of the theorem say when $f(x) = \cos \alpha x$?

DERIVATIVE THEOREM

If $f(x)$ has the Fourier transform $F(s)$ then $f'(x)$ has the Fourier transform $i2\pi sF(s)$.

Derivation:

$$\int_{-\infty}^{\infty} f'(x)e^{-i2\pi xs}\,dx = \int_{-\infty}^{\infty} \lim \frac{f(x+\Delta x) - f(x)}{\Delta x} e^{-i2\pi xs}\,dx$$
$$= \lim \int_{-\infty}^{\infty} \frac{f(x+\Delta x)}{\Delta x} e^{-i2\pi xs}\,dx - \lim \int_{-\infty}^{\infty} \frac{f(x)}{\Delta x} e^{-i2\pi xs}\,dx$$

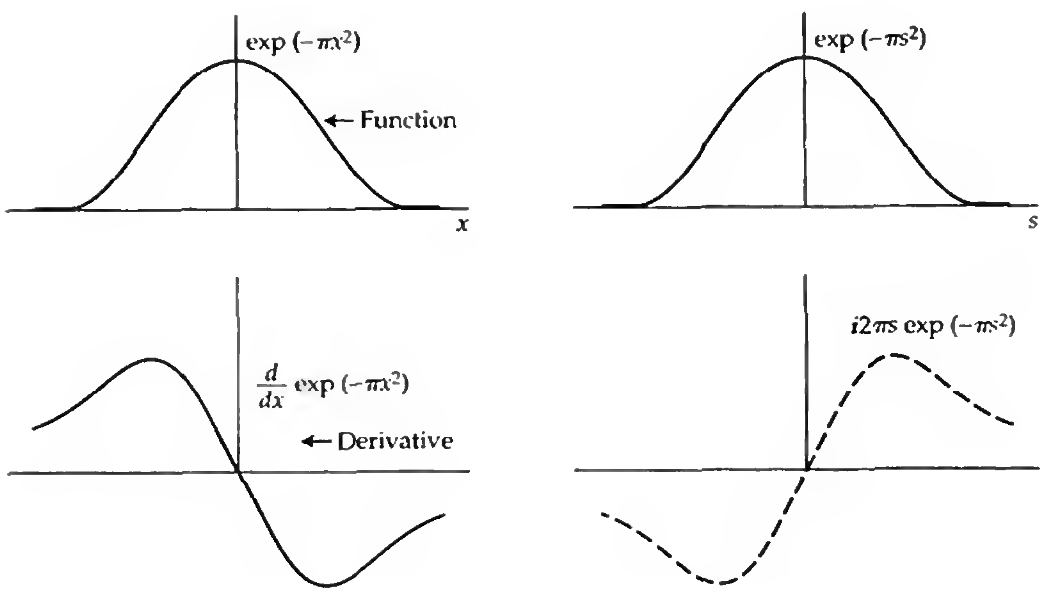


Fig. 6.13 **Differentiation of a function incurs multiplication of the transform by** $i2\pi s$.

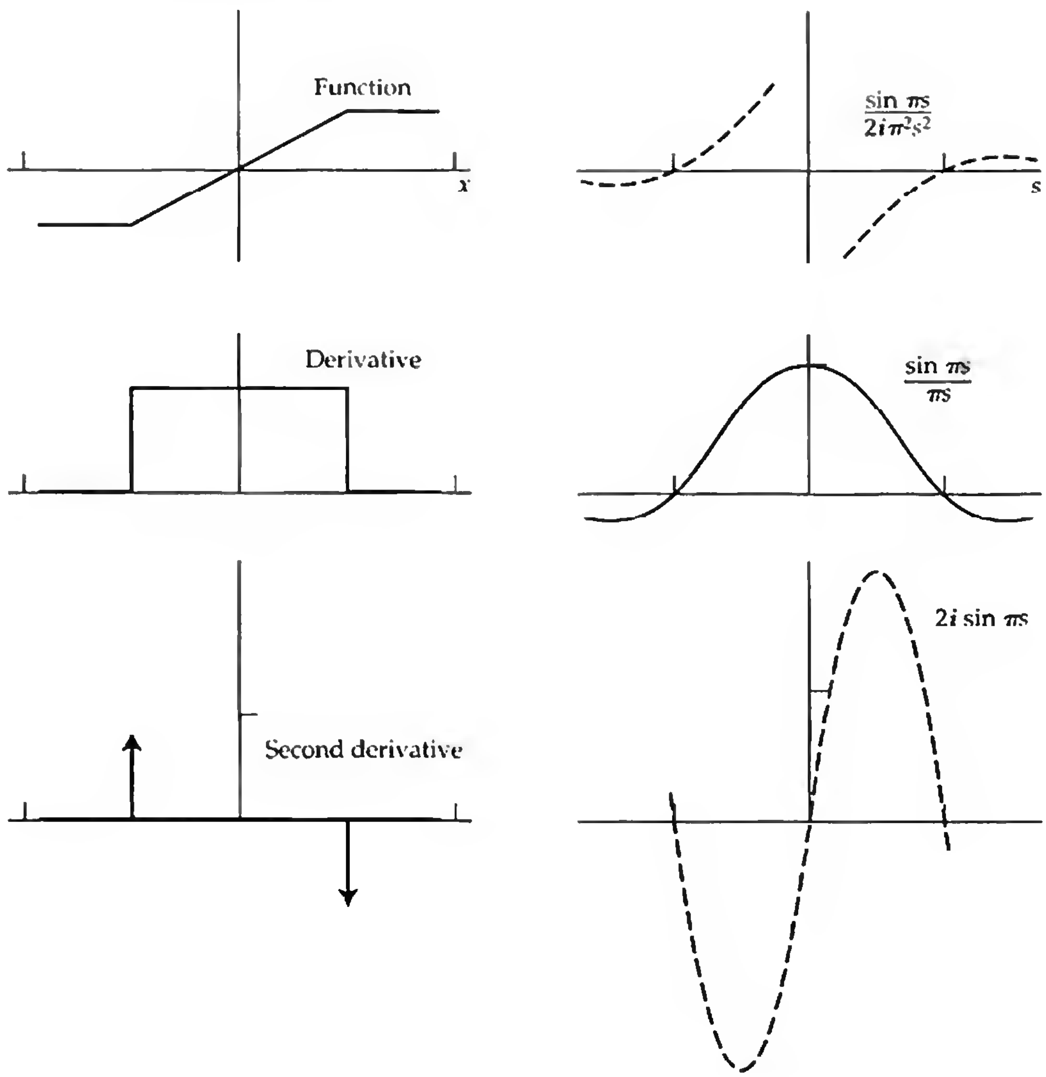


Fig. 6.14 **Successive applications of the derivative theorem.**

$$= \lim \frac{e^{i2\pi\,\Delta x s}F(s) - F(s)}{\Delta x}$$
$$= i2\pi s F(s).$$

Since taking the derivative of a function multiplies its transform by $i2\pi s$, we can say that differentiation enhances the higher frequencies, attenuates the lower frequencies, and suppresses any zero-frequency component. Examples are given in Figs. 6.13 and 6.14.

It happens quite frequently that the multiplication of $i2\pi s$ causes the integral of $|i2\pi s F(s)|$ to diverge. Correspondingly, the derivative $f'(x)$ will exhibit infinite discontinuities. Such situations are accommodated by the impulse symbol and its derivatives.

DERIVATIVE OF A CONVOLUTION INTEGRAL

From the derivative theorem taken in conjunction with the convolution theorem, it follows that if

$$h = f * g,$$

then
$$h' = f' * g,$$

and also
$$h' = f * g'.$$

Derivation:

$$\begin{aligned}\frac{d}{dx}[f(x) * g(x)] &\supset i2\pi s[F(s)G(s)]\\ f'(x) * g(x) &\supset [i2\pi sF(s)]G(s)\\ f(x) * g'(x) &\supset F(s)[i2\pi sG(s)].\end{aligned}$$

These conclusions may be stated in a different form as follows:

The derivative of a convolution is the convolution of either of the functions with the derivative of the other.

Thus

$$(f * g)' = f' * g = f * g',$$

and again
$$\begin{aligned}f * g &= f' * \int^x g\,dx\\ &= f'' * \int\int^x g\,dx\,dx\\ &\quad \ldots .\end{aligned}$$

Exercise. Investigate the question of what lower limits are appropriate for the integrals. Investigate the formula $f' * g = f * g'$ by integration by parts.

In terms of the derivative of the impulse symbol we may write

$$h = \delta * h.$$

Therefore
$$\begin{aligned}h' &= \delta' * h = \delta' * f * g = (\delta' * f) * g = f' * g\\ &= f * (\delta' * g) = f * g',\\ f * g &= \delta * f * g = (\delta' * H) * f * g = (\delta' * f) * (H * g).\end{aligned}$$

The formulas quoted here are applicable to the evaluation of particular convolution integrals analytically or numerically but are principally of theoretical value (for example, in the deduction of the uncertainty relation and in deriving the formulas for the response of a filter in terms of its impulse and step responses).

The convenient algebra permitted by the δ notation in conjunction with the associative and other properties of convolution enables rapid generation of rela-

tions which may be needed for some problem under study. Many interesting possibilities arise. For example, starting from

$$\frac{d}{dx}(f * g) \supset i2\pi sFG$$

we may factor the right-hand side to get

$$\frac{d}{dx}(f * g) = \frac{d^{\frac{1}{2}}f}{dx^{\frac{1}{2}}} * \frac{d^{\frac{1}{2}}g}{dx^{\frac{1}{2}}} \supset (i2\pi s)^{\frac{1}{2}}F(i2\pi s)^{\frac{1}{2}}G.$$

■

THE TRANSFORM OF A GENERALIZED FUNCTION

Let $p(x)$ be a generalized function defined as in Chapter 5 by the sequence $p_\tau(x)$. Let the Fourier transforms of members of the defining sequence be $P_\tau(s)$, where

$$P_\tau(s) = \int_{-\infty}^{\infty} p_\tau(x)e^{-i2\pi sx}\,dx.$$

Perhaps this new sequence $P_\tau(s)$ defines a generalized function. We know that the members of the sequence are particularly well-behaved; we test the sequence for regularity by means of an arbitrary particularly well-behaved function $F(x)$ whose Fourier transform is $\overline{F}(s)$. From the energy theorem

$$\int_{-\infty}^{\infty} P_\tau(s)\overline{F}(s)\,ds = \int_{-\infty}^{\infty} p_\tau(x)F(-x)\,dx$$

it follows that

$$\lim_{\tau\to 0}\int_{-\infty}^{\infty} P_\tau(s)\overline{F}(s)\,ds = \lim_{\tau\to 0}\int_{-\infty}^{\infty} p_\tau(x)F(-x)\,dx,$$

and we know the latter limit exists. Hence $P_\tau(s)$ defines a generalized function, to which we give the symbol $P(s)$ and the meaning "Fourier transform of the generalized function $p(x)$."

The following statement can be made about $p(x)$ and $P(s)$,

$$\int_{-\infty}^{\infty} P(s)\overline{F}(s)\,ds = \int_{-\infty}^{\infty} p(x)F(-x)\,ds,$$

where $F(x)$ is any particularly well-behaved function and $\overline{F}(s)$ is its Fourier transform.

Let $\phi(x)$ be a function, such as a polynomial, that has derivatives of all orders at all points but whose behavior as $|x| \to \infty$ is not so stringently controlled as that of particularly well-behaved functions. We allow $\phi(x)$ to go infinite as $|x|^N$ where N is finite. Functions such as exp x and log x would be excluded. Then products of $\phi(x)$ and any particularly well-behaved function will eventually be over-

whelmed by the latter as $|x| \to \infty$, since the particularly well-behaved factor dies out faster than $|x|^{-N}$. Furthermore, since the product has all derivatives at all points, the product itself is particularly well-behaved.

Consider a sequence

$$\phi(x)p_\tau(x).$$

It is particularly well-behaved, and

$$\lim_{\tau\to 0}\int_{-\infty}^{\infty}[\phi(x)p_\tau(x)]F(x)\,dx = \lim_{\tau\to 0}\int_{-\infty}^{\infty}p_\tau(x)[\phi(x)F(x)]\,dx,$$

which exists since $p_\tau(x)$ is a regular sequence and $[\phi(x)F(x)]$ is particularly well-behaved. The generalized function so defined we write as

$$\phi(x)p(x).$$

In practice we use this notation with functions $\phi(x)$ that have a *sufficient* number of derivatives and include exponentially increasing functions when, as is the case with the pulse sequence $\tau^{-1}\Pi(x/\tau)$, the behavior at infinity is inessential.

Nothing is introduced which could be called the product of two generalized functions; the product of two defining sequences is not necessarily a regular sequence and consequently does not in general define a generalized function.

PROOFS OF THEOREMS

The numerous theorems of Fourier theory, which have proved so fruitful in the preceding sections, have shown themselves perfectly adaptable to the insertion of the impulse symbol $\delta(x)$, the shah symbol $\text{III}(x)$, the duplicating symbol $\text{II}(x)$, and other familiar nonfunctions. Standard proofs of the theorems eliminated these cases, and accepted conditions for the applicability of the theorems which we have found in practice need not be observed. This situation was dealt with above by introducing the idea of a transform in the limit, and special ad hoc interpretations as limits were placed on expressions containing impulse symbols.

Having established the algebra for generalized functions, we can also give systematic proofs of the various theorems, free from the awkward conditions that arise when attention is confined to those functions that meet the conditions for existence of regular transforms. The difficulties associated with functions that do not have derivatives disappear, for generalized functions possess derivatives of all orders. And more intolerable circumstances, such as the lack of a regular spectrum for direct current, also vanish. The brief "derivations" given above are useful while the theorems are being learned but do not qualify as strict proofs as presented by Lighthill (1958). Some examples of proofs in terms of the sequences $p_\tau(x)$ used for defining generalized functions are given below as illustrations of formal proofs.

Similarity and shift theorems. We prove these theorems simultaneously. Let $p(x)$ be a generalized function with Fourier transform $P(s)$. Then

$$p(ax + b) \supset \frac{1}{|a|} e^{i2\pi bs/a} P\left(\frac{s}{a}\right).$$

Proof: Since

$$\lim_{\tau \to 0} \int_{-\infty}^{\infty} p_\tau(ax + b)F(x)\, dx = \frac{1}{|a|} \lim_{\tau \to 0} \int_{-\infty}^{\infty} p_\tau(x)F\left(\frac{x - b}{a}\right) dx$$

exists, we have a meaning for $p(ax + b)$. Now

$$p_\tau(ax + b) \supset \frac{1}{|a|} e^{i2\pi bs/a} P_\tau\left(\frac{s}{a}\right)$$

by substitution of variables. Hence the two theorems follow.

Derivative theorem. The Fourier transform of $p'_\tau(x)$ is $i2\pi s P_\tau(s)$. Hence the Fourier transform of $p'(x)$ is $i2\pi s P(s)$.

Power theorem. Since no meaning has been assigned to the product of two generalized functions, the best theorem that can be proved is

$$\int_{-\infty}^{\infty} P(s)\overline{F}(s)\, ds = \int_{-\infty}^{\infty} p(x)F(-x)\, dx,$$

where $F(x)$ is a particularly well-behaved function and $p(x)$ is a generalized function. The theorem follows from the fact that

$$\lim_{\tau \to 0} \int_{-\infty}^{\infty} P_\tau(s)\overline{F}(s)\, ds = \lim_{\tau \to 0} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \overline{F}(s)p(x)e^{-i2\pi sx}\, dx\, ds$$
$$= \lim_{\tau \to 0} \int_{-\infty}^{\infty} p_\tau(x)F(-x)\, dx.$$

All the above theorems generalize readily to two dimensions. Rotation and shear, which do not arise in one dimension, are associated with new basic theorems, circular symmetry introduces important special cases, and the affine theorem applying to functions of the form $f(ax + by + c, dx + ey + f)$ has rich implications for graphics (Bracewell et al. 1993, Bracewell 1994).

SUMMARY OF THEOREMS

The theorems discussed in the preceding pages are collected for reference in Table 6.1.

TABLE 6.1
Theorems for the Fourier transform

Theorem	$f(x)$	$F(s)$
Similarity	$f(ax)$	$\frac{1}{\lvert a\rvert}F\left(\frac{s}{a}\right)$
Addition	$f(x) + g(x)$	$F(s) + G(s)$
Shift	$f(x - a)$	$e^{-i2\pi as}F(s)$
Modulation	$f(x)\cos\omega x$	$\frac{1}{2}F\left(s - \frac{\omega}{2\pi}\right) + \frac{1}{2}F\left(s + \frac{\omega}{2\pi}\right)$
Convolution	$f(x) * g(x)$	$F(s)G(s)$
Autocorrelation	$f(x) * f^*(-x)$	$\lvert F(s)\rvert^2$
Derivative	$f'(x)$	$i2\pi sF(s)$
Derivative of convolution	$\frac{d}{dx}[f(x) * g(x)] = f'(x) * g(x) = f(x) * g'(x)$	
Rayleigh	$\int_{-\infty}^{\infty} \lvert f(x)\rvert^2\,dx = \int_{-\infty}^{\infty} \lvert F(s)\rvert^2\,ds$	
Power	$\int_{-\infty}^{\infty} f(x)g^*(x)\,dx = \int_{-\infty}^{\infty} F(s)G^*(s)\,ds$	
(f and g real)	$\int_{-\infty}^{\infty} f(x)g(-x)\,dx = \int_{-\infty}^{\infty} F(s)G(s)\,ds$	

BIBLIOGRAPHY

Bell, R. J.: "Introductory Fourier Transform Spectroscopy," Academic Press, New York, 1972.

Bracewell, R. N.: "Affine Theorem for the Hartley Transform of an Image," *Proc. IEEE*, vol. 82, pp. 388–390, 1994.

Bracewell, R. N., K. Y. Chang, A. K. Jha, and Y. H. Wang: "Affine Theorem for Two-dimensional Fourier Transform," *Electronics Letters*, vol. 29, p. 304, 1993.

Lighthill, M. J.: "An Introduction to Fourier Analysis and Generalized Functions," Cambridge University Press, Cambridge, England, 1958.

Lord Rayleigh: "On the Character of the Complete Radiation at a Given Temperature," *Phil. Mag.* ser. 5, vol. 27, 1889, *Scientific Papers*, Cambridge University Press, Cambridge, England, 1902, and Dover Publications, New York, vol. 3, p. 273, 1964.

Titchmarsh, E. C.: "A Contribution to the Theory of Fourier Transforms," *Proc. Lond. Math. Soc.*, vol. 23, p. 279, 1924.

Wiener, N.: "Extrapolation, Interpolation, and Smoothing of Stationary Time Series," John Wiley and Sons, New York, 1949.

PROBLEMS

1. Using the transform pairs given for reference, deduce the further pairs listed below by application of the appropriate theorem. Assume that A and σ are positive.

$$\frac{\sin x}{x} \supset \pi\Pi(\pi s) \qquad \left(\frac{\sin x}{x}\right)^2 \supset \pi\Lambda(\pi s)$$

$$\frac{\sin Ax}{Ax} \supset \frac{\pi}{A}\Pi\left(\frac{\pi s}{A}\right) \qquad \left(\frac{\sin Ax}{Ax}\right)^2 \supset \frac{\pi}{A}\Lambda\left(\frac{\pi s}{A}\right)$$

$$e^{-x^2} \supset \pi^{\frac{1}{2}}e^{-\pi^2 s^2} \qquad \delta(ax) \supset \frac{1}{|a|}$$

$$e^{-Ax^2} \supset \left(\frac{\pi}{A}\right)^{\frac{1}{2}}e^{-\pi^2 s^2/A} \qquad \delta(ax+b) \supset \frac{1}{|a|}e^{i2\pi bs/a}$$

$$e^{-x^2/2\sigma^2} \supset (2\pi)^{\frac{1}{2}}\sigma e^{-2\pi^2\sigma^2 s^2} \qquad e^{ix} \supset \delta\left(s - \frac{1}{2\pi}\right)$$

2. Show that the following transform pairs follow from the addition theorem, and make graphs.

$$1 + \cos \pi x \supset \delta(s) + \mathrm{II}(s)$$
$$1 + \sin \pi x \supset \delta(s) + i\mathrm{I}_{\mathrm{I}}(s)$$
$$\operatorname{sinc} x + \tfrac{1}{2}\operatorname{sinc}^2 \tfrac{1}{2}x \supset \Pi(s) + \Lambda(2s)$$
$$A^{-\frac{1}{2}}e^{-\pi x^2/A} + A^{\frac{1}{2}}e^{-\pi Ax^2} \supset e^{-\pi s^2/A} + e^{-\pi As^2}$$
$$4\cos^2 \pi x + 4\cos^2 \tfrac{1}{2}\pi x - 3 \supset \delta(s+\tfrac{1}{2}) + \delta(s-\tfrac{1}{2}) + \delta(s) + \tfrac{1}{2}\delta(s-1) + \tfrac{1}{2}\delta(s+1).$$

3. Deduce the following transform pairs, using the shift theorem.

$$\frac{\cos \pi x}{\pi(x-\frac{1}{2})} \supset -e^{-i\pi s}\Pi(s)$$
$$\frac{\sin \pi x}{\pi(x-1)} \supset -e^{-i2\pi s}\Pi(s)$$
$$\Lambda(x-1) \supset e^{-i2\pi s}\operatorname{sinc}^2 s$$
$$\Pi(x-\tfrac{1}{2}) \supset e^{-i\pi s}\operatorname{sinc} s$$
$$\Pi(x)\operatorname{sgn} x \supset -i\sin \tfrac{1}{2}\pi s \operatorname{sinc} \tfrac{1}{2}s$$
$$\Pi\left(\frac{x-\frac{1}{2}a}{a}\right) \supset |a|e^{-i\pi as}\operatorname{sinc} as.$$

4. Use the convolution theorem to find and graph the transforms of the following functions: $\operatorname{sinc} x \operatorname{sinc} 2x$, $(\operatorname{sinc} x \cos 10x)^2$.

5. Let $f(x)$ be a periodic function with period a, that is, $f(x+a) = f(x)$ for all x. Since the Fourier transform of $f(x+a)$ is, by the shift theorem, equal to $\exp(i2\pi as)F(s)$, which must be equal to $F(s)$, what can be deduced about the transform of a periodic function?

6. Graph the transform of $f(x)\sin \omega x$ for large and small values of ω, and explain graphically how, for small values of ω, the transform of $f(x)\sin \omega x$ is proportional to the derivative of the transform of $f(x)$. ▷

7. Graph the transform of $\exp(-x)H(x)\cos \omega x$. Is it an even function of s?

8. Show that a pulse signal described by $\Pi(x/X)\cos 2\pi fx$ has a spectrum

$$\tfrac{1}{2}X\{\operatorname{sinc}[X(s+f)] + \operatorname{sinc}[X(s-f)]\}$$

9. Show that a modulated pulse described by $\Pi(x/X)(1 + M\cos 2\pi Fx)\cos 2\pi fx$ has a spectrum

$$\tfrac{1}{2}X\{\operatorname{sinc}[X(s+f)] + \operatorname{sinc}[X(s-f)]\} + \tfrac{1}{4}MX\{\operatorname{sinc}[X(s+f+F)] \\ + \operatorname{sinc}[X(s+f-F)] + \operatorname{sinc}[X(s-f+F)] + \operatorname{sinc}[X(s-f-F)]\}.$$

Graph the spectrum to a suitably exaggerated scale for a case where there are 100 modulation cycles and 100,000 radio-frequency cycles in one pulse and the modulation coefficient M is 0.6. Show by dimensioning how the factors 100, 100,000, and 0.6 enter into the shape of the spectrum.

10. A function $f(x)$ is defined by

$$f(x) = \begin{cases} 0 & |x| > 2 \\ 2 - |x| & 1 < |x| < 2 \\ 1 & |x| < 1; \end{cases}$$

show that

$$f(x) = 2\Lambda\left(\frac{x}{2}\right) - \Lambda(x) = \Lambda(x) * [\delta(x+1) + \delta(x) + \delta(x-1)]$$

and hence that

$$F(s) = 4\operatorname{sinc}^2 2s - \operatorname{sinc}^2 s = \operatorname{sinc}^2 s(1 + 2\cos 2\pi s).$$

11. Prove that $f * g * h \supset FGH$ and hence that $f^{*n} \supset F^n$.

12. The notation f^{*n} meaning $f(x)$ convolved with itself $n-1$ times, where $n = 2, 3, 4, \ldots,$ suggests the idea of fractional-order self-convolution. Show that such a generalization of convolution is readily made and that, for example, one reasonable expression for $f(x)$ convolved with itself half a time would be

$$f^{*1\frac{1}{2}} \equiv \int e^{i2\pi sx}\left[\int e^{-i2\pi su} f(u)\,du\right]^{1\frac{1}{2}} ds. \triangleright$$

13. Prove that

$$(f * g)(h * j) \supset (FG) * (HJ)$$

and that

$$(f + g) * (h + j) \supset FH + FJ + GH + GJ.$$

14. Use the convolution theorem to obtain an expression for

$$e^{-ax^2} * e^{-bx^2}. \triangleright$$

15. Prove that

$$\int_{-\infty}^{\infty} f^*(u) g^*(x-u)\,du \supset F^*(-s)G^*(-s).$$

16. Prove that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f^*(u) g^*(u-x) e^{-i2\pi xs}\,du\,dx = F^*(-s)G^*(s).$$

17. Show by Rayleigh's theorem that

$$\int_{-\infty}^{\infty} \operatorname{sinc}^2 x \, dx = 1$$

$$\int_{-\infty}^{\infty} \operatorname{sinc}^4 x \, dx = \int_{-\infty}^{\infty} [\Lambda(x)]^2 \, dx = \tfrac{2}{3}$$

$$\int_{-\infty}^{\infty} [J_0(x)]^2 dx = \infty$$

$$\int_{-\infty}^{\infty} \frac{dx}{(1 + x^2)^2} = \frac{\pi}{2}$$

18. Complete the following schemata for reference, including thumbnail sketches of the functions.

function $\supset$	transform
autocorrelation $f \star f$ $\supset$	power spectrum $\lvert F(s)\rvert^2$

$\Pi(x)$	

$\Lambda(x)$	

$\gamma(x)$	

$\operatorname{sinc} x$	

$\Pi(x) \cos 2\pi f x$	

19. Show the fallacy in the following reasoning. "The Fourier transform of $\int_{-\infty}^{x} f(x)\,dx$ must be $F(s)/i2\pi s$ because the derivative of $\int_{-\infty}^{x} f(x)\,dx$ is $f(x)$, and hence by the derivative theorem the transform of $f(x)$ would be $F(s)$, which is true."

20. Establish an integral theorem for the Fourier transform of the indefinite integral of a function. ▷

21. Use the derivative theorem to find the Fourier transform of $xe^{-\pi x^2}$.

22. Show that $2\pi x \Pi(x) \supset i \operatorname{sinc}' s$.

23. The following brief derivation appears to show that the area under a derivative is zero. Thus

$$\int_{-\infty}^{\infty} f'(x)\,dx = \overline{f'(x)}\Big|_0 = i2\pi s F(s)\Big|_0 = 0.$$

Confirm that this is so, or find the error in reasoning.

24. Show that

$$f(ax - b) \supset \frac{1}{|a|} e^{-i2\pi bs/a} F\left(\frac{s}{a}\right).$$

25. Show from the energy theorem that

$$\int_{-\infty}^{\infty} e^{-\pi x^2} \cos 2\pi ax \, dx = e^{-\pi a^2}.$$

26. Show from the energy theorem that

$$\int_{-\infty}^{\infty} \operatorname{sinc}^2 x \cos \pi x \, dx = \tfrac{1}{2}. \triangleright$$

27. Show that the function whose Fourier transform is $|\operatorname{sinc} s|$ has a triangular autocorrelation function.

28. As a rule, the autocorrelation function tends to be more spread out than the function it comes from. But show that

$$\frac{1}{\pi x} * \frac{-1}{\pi x} = \delta(x).$$

Show that $(\pi x)^{-1}$ must have a flat energy spectrum, and from that deduce and investigate other functions whose autocorrelation is impulsive. ▷

29. The Maclaurin series for $F(s)$ is

$$F(0) + sF'(0) + \frac{s^2}{2!} F''(0) + \dots .$$

Consider the case of $F(s) = \exp(-\pi s^2)$, where the series is known to converge and to converge to $F(s)$. Thus, in this particular case,

$$F(s) = \sum_{n=0}^{\infty} \frac{s^n}{n!} F^{(n)}(0).$$

If $F(s)$ is the transform of $F(x)$, then transforming this equation we obtain

$$f(x) = \delta(x) \int_{-\infty}^{\infty} f(x)\, dx - \delta'(x) \int_{-\infty}^{\infty} x f(x)\, dx + \delta''(x) \int_{-\infty}^{\infty} \frac{x^2}{2!} f(x)\, dx + \dots .$$

How do you explain this result? ▷

30. Fourier pairs. Derive the following Fourier transform pairs:

(*a*) $e^{i\pi x^2} \supset e^{i\pi/4} e^{-i\pi s^2}$

(*b*) $\cos(\pi x^2) \supset 2^{-\frac{1}{2}}[\cos(\pi s^2) + \sin(\pi s^2)]$

(*c*) $\sin(\pi x^2) \supset 2^{-\frac{1}{2}}[\cos(\pi s^2) - \sin(\pi s^2)]$

(*d*) $e^{-\pi\alpha x^2} \begin{matrix}\cos\\ \sin\end{matrix} \pi\beta x^2 \supset (\alpha^2+\beta^2)^{-\frac{1}{4}} \exp\left(-\dfrac{\pi\alpha s^2}{\alpha^2+\beta^2}\right) \begin{matrix}\cos\\ \sin\end{matrix} \left(\arctan\left(\dfrac{\beta}{\alpha}\right) - \dfrac{\pi\beta s^2}{\alpha^2+\beta^2}\right)$

(*e*) $e^{-\pi(\alpha+i\beta)x^2} \supset (\alpha+i\beta)^{-\frac{1}{2}} e^{-\pi s^2/(\alpha+i\beta)}$

31. Frequency analysis. A volcano on the floor of the Pacific Ocean erupted near an inhabited island, causing the sea surface to rise and fall, reaching a maximum height of about 10 meters. The height was recorded by the captain of a vessel standing offshore, using a sextant to determine the distance from the water to the top of a cliff. Later examination showed that the height $h(t)$ could be represented approximately by

$$h(t) = 11\sin(45° - 72°t)\exp(-t^2/5),$$

where h is in meters and t is in minutes. The volcano erupts from time to time, often causing damage to the docks and shipping in the lagoon, but this is the first time a waveform has become available and it is to be the basis of a redesign of the port.

(*a*) Paying particular attention to the correctness of numerical values, but not necessarily carrying out all the arithmetic, obtain the Fourier transform $\overline{h}(f)$ of $h(t)$.

(*b*) At what frequency, in cycles per minute, will the excitation be at a maximum?

32. Chirp signal. A chirp is a signal that sweeps in frequency and is used in radar by bats and humans to facilitate the sorting out of the emitted signal from the echo under conditions where the first echoes will be returning while the emission is still continuing. An example is

$$s(t) = e^{-\pi t^2/T^2} e^{i2\pi(f_0 t + \beta t^2)}.$$

This chirped pulse has an equivalent duration T, a frequency f_0 at midpulse, and a frequency sweep rate 2β. Show that the power spectrum is centered at f_0 and has an equivalent width Δ given by $\Delta = 2^{-\frac{1}{2}}T^{-1}(1 + 4\beta^2T^4)^{\frac{1}{2}}$.

33. Voigt profiles. Spectral lines often have a profile of the form $[1 + 4(f - f_0)^2/B^2]^{-1}$, which arises, for example, from absorption by a resonator. This is a shifted Cauchy profile. Other spectral lines may have a Gaussian profile, as in the case of a gas in a state of turbulence where Doppler shifts greatly exceed the natural linewidth B. Intermediate profiles of the form Gaussian-convolved-with-Cauchy are known as Voigt profiles. They have interesting properties. Show that the convolution of two Voigt profiles is also a Voigt profile. ▷

34. Inverse theorems. Show that the inverse derivative and inverse shift theorems are

$$-i2\pi x f(x) \supset F'(s)$$
$$e^{i2\pi s_0 x} f(x) \supset F(s - s_0).$$

Are there any other theorems where the direct and inverse forms are not the same?

7

Obtaining Transforms

A number of transforms were introduced earlier to illustrate the basic theorems of the Fourier transformation. Of course, one need not necessarily be aware of any particular Fourier transform pairs to appreciate the meaning of the theorems. Many of the chains of argument in which the Fourier transformation is important are independent of any knowledge of particular examples. Even so, carrying out a general argument with a special case in mind often serves as insurance against surprises.

The examples of Fourier transform pairs chosen for illustration were all introduced without derivation and asserted to be verifiable by evaluation of the Fourier integral. Obviously this does not help when it is necessary to generate new pairs. We therefore consider various ways of carrying out the Fourier transformation. Numerical methods based on the discrete Fourier transform and allowing for use of the fast Fourier transform algorithm are discussed in Chapter 11, but this chapter also advocates the slow Fourier transform algorithm that simply evaluates the integral as a sum. The program is short, completely under the user's control, and runs in a few seconds for the majority of everyday applications.

Starting from the given function $f(x)$ whose Fourier transform is to be deduced, one may first contemplate the integral

$$\int_{-\infty}^{\infty} f(x)e^{-i2\pi xs}\,dx.$$

If this integral can be evaluated for all s, the problem is solved.

A number of different approaches from the standpoint of integration are discussed separately in subsequent sections.

In addition to this direct approach we have a powerful resource in the basic theorems which have now been established. Many of the theorems take the form "If f and F are a transform pair then g and G are also." Thus, if one knows a trans-

form pair to begin with, others may be generated. It is indeed possible to build up an extensive dictionary of transforms by means of the theorems, starting from the beginnings already laid down. It may be surmised that some classes of function will never be stumbled on in this way; on the other hand, a variety of physically feasible functions prove to be accessible. Generations from theorems is taken up later with examples. Finally, there is the possibility of extracting a desired transform from tables.

INTEGRATION IN CLOSED FORM

It is a propitious circumstance if $f(x)$ is zero over some range of x and if in addition its behavior is simple where it is nonzero. Thus if

$$f(x) = \Pi(x),$$

then

$$\begin{aligned}\int_{-\infty}^{\infty} f(x)e^{-i2\pi xs}\,dx &= \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{-i2\pi xs}\,dx \\ &= \int_{-\frac{1}{2}}^{\frac{1}{2}} \cos 2\pi xs\,dx \\ &= \frac{\sin \pi s}{\pi s} \\ &= \operatorname{sinc} s.\end{aligned}$$

If $f(x)$ is nonzero on several segments of the abscissa and constant within each segment, the integration in closed form can likewise be done. Thus if we let

$$f(x) = a\Pi\left(\frac{x-b}{c}\right),$$

then

$$\begin{aligned}F(s) &= \int_{b-\frac{1}{2}c}^{b+\frac{1}{2}c} ae^{-i2\pi xs}\,dx \\ &= a\int_{-\frac{1}{2}c}^{\frac{1}{2}c} e^{-i2\pi(u+b)s}\,du \\ &= ae^{-i2\pi bs}\int_{-\frac{1}{2}c}^{\frac{1}{2}c} \cos 2\pi us\,du \\ &= ace^{-i2\pi bs}\operatorname{sinc} cs.\end{aligned}$$

It follows that if

$$f(x) = \sum_n a_n\Pi\left(\frac{x-b_n}{c_n}\right),$$

then

$$F(s) = \sum_n a_n c_n e^{-i2\pi b_n s}\operatorname{sinc} c_n s.$$

Functions built up segmentally of rectangle functions include staircase functions and functions suitable for discussing Morse code, teleprinter signals, and on/off servomechanisms. They not only occur frequently in engineering but can also simulate, as closely as may be desired, any kind of variation.

An even simpler case arises if $f(x)$ is zero almost everywhere; for example, suppose that $f(x)$ is a set of impulses of various strengths at various values of x:

$$f(x) = \sum_n a_n\,\delta(x - b_n).$$

Then only the values of $a_n \exp(-i2\pi xs)$ at the points $x = b_n$ can matter. By the sifting theorem for the impulse symbol,

$$\begin{aligned}
F(s) &= \int_{-\infty}^{\infty} \sum a_n\,\delta(x - b_n)e^{-i2\pi xs}\,dx \\
&= \int_{b_1-}^{b_1+} a_1\,\delta(x - b_1)e^{-i2\pi xs}\,dx + \int_{b_2-}^{b_2+} a_2\,\delta(x - b_2)e^{-i2\pi xs}\,dx + \dots \\
&= a_1 e^{-i2\pi b_1 s} + a_2 e^{-i2\pi b_2 s} + \dots .
\end{aligned}$$

If $f(x)$ has some special functional form, it may prove possible to perform the integration, but no general rules can be given. Some simple examples follow.

1. Let $f(x) = \operatorname{sinc} x$. Then

$$\begin{aligned}
F(s) &= \int_{-\infty}^{\infty} \operatorname{sinc} x\, e^{-i2\pi xs}\,dx \\
&= \int_{-\infty}^{\infty} \frac{\sin \pi x \cos 2\pi xs}{\pi x}\,dx \\
&= \int_{-\infty}^{\infty} \left[\frac{\sin(\pi x + 2\pi xs)}{2\pi x} + \frac{\sin(\pi x - 2\pi xs)}{2\pi x}\right] dx \\
&= \int_{-\infty}^{\infty} \left\{\frac{1 + 2s}{2}\operatorname{sinc}[(1 + 2s)x] + \frac{1 - 2s}{2}\operatorname{sinc}[(1 - 2s)x]\right\} dx \\
&= \frac{1 + 2s}{2|1 + 2s|} + \frac{1 - 2s}{2|1 - 2s|} \\
&= \Pi(s).
\end{aligned}$$

We have used the result that

$$\int_{-\infty}^{\infty} \operatorname{sinc} ax\,dx = \frac{1}{|a|}.$$

2. Let $f(x) = e^{-|x|}$. Then

$$\begin{aligned}
F(s) &= \int_{-\infty}^{\infty} e^{-|x|}e^{-i2\pi xs}\,dx \\
&= 2\int_0^{\infty} e^{-x} \cos 2\pi xs\,dx
\end{aligned}$$

$$= 2\,\mathrm{Re}\int_0^\infty e^{-x}e^{i2\pi xs}\,dx$$

$$= 2\,\mathrm{Re}\int_0^\infty e^{(i2\pi s-1)x}\,dx$$

$$= 2\,\mathrm{Re}\,\frac{-1}{i2\pi s - 1}$$

$$= \frac{2}{4\pi^2 s^2 + 1}.$$

3. Let $f(x) = e^{-\pi x^2}$. Then

$$F(s) = \int_{-\infty}^{\infty} e^{-\pi x^2} e^{-i2\pi xs}\,dx$$

$$= \int_{-\infty}^{\infty} e^{-\pi(x^2 + i2xs)}\,dx$$

$$= e^{-\pi s^2}\int_{-\infty}^{\infty} e^{-\pi(x+is)^2}\,dx$$

$$= e^{-\pi s^2}\int_{-\infty}^{\infty} e^{-\pi(x+is)^2}\,d(x+is)$$

$$= e^{-\pi s^2}.$$

In this example, we have used the known result that the infinite integral of $\exp(-\pi x^2)$ is unity. The next case illustrates a contour integral.

4. Let $f(x) = x^{-1}$. The infinite discontinuity at the origin causes the standard Fourier integral to diverge; hence we consider instead

$$\lim_{\epsilon\to 0}\left(\int_{-\infty}^{-\epsilon} \frac{e^{-i2\pi sx}}{x}\,dx + \int_{\epsilon}^{\infty}\right).$$

Now consider

$$\int_C \frac{e^{-i2\pi sz}}{z}\,dz,$$

where the contour C is a semicircle of radius R in the complex plane of z, whose diameter lies along the real axis and has a small indentation of radius ϵ at the origin. The contour integral is zero since no poles are enclosed. Thus

$$\int_{-R}^{-\epsilon}\frac{e^{-i2\pi sx}}{x}\,dx + \int_{\pi}^{0} ie^{-i2\pi s\epsilon e^{i\theta}}\,d\theta + \int_{\epsilon}^{R}\frac{e^{-i2\pi sx}}{x}\,dx + \int_0^{\pi} ie^{-i2\pi sRe^{i\theta}}\,d\theta = 0.$$

As $R\to\infty$, the fourth integral vanishes and the second equals $\pm i\pi$ according to the sign of s, so that we have

$$\lim_{\epsilon\to 0}\left(\int_{-\infty}^{-\epsilon}\frac{e^{-i2\pi sx}}{x}\,dx + \int_{\epsilon}^{\infty}\right) + i\pi\,\mathrm{sgn}\,s = 0.$$

Hence the desired transform pair in the limit is

$$\frac{1}{x} \supset -i\pi\,\mathrm{sgn}\,s.$$

5. Let $f(x) = \operatorname{sgn} x$. This is the previous example in reverse, and it is seen that in this case the Fourier integral fails to exist in the standard sense because the function does not possess an absolutely convergent integral. Therefore consider a sequence of transformable functions which approach sgn x as a limit, for example, the sequence $\exp(-r|x|) \operatorname{sgn} x$ as $\tau \to 0$. The transform will be

$$\int_{-\infty}^{\infty} e^{-\tau|x|} \operatorname{sgn} x \, e^{-i2\pi xs} \, dx = \int_{-\infty}^{0} -e^{(\tau - i2\pi s)x} \, dx + \int_{0}^{\infty} e^{-(\tau + i2\pi s)x} \, dx$$
$$= -\frac{1}{\tau - i2\pi s} + \frac{1}{\tau + i2\pi s}.$$

As $\tau \to 0$ this expression has a limit $1/i\pi s$. Hence we have the Fourier transform pair in the limit,

$$\operatorname{sgn} x \supset \frac{1}{i\pi s}.$$

NUMERICAL FOURIER TRANSFORMATION

If the values of a function have been obtained by physical measurement and it is necessary to find its Fourier transform, various possibilities exist.

First of all, certain limitations of physical data will influence the result, and we shall begin with this aspect of numerical transformation.

The data will be given at discrete values of the independent variable x. The interval Δx may be so fine that there is no concern about interpolating intermediate values, but in any case we can take the view that the data can hardly contain significant information about Fourier components with periods less than $2\,\Delta x$. Therefore it is not necessary to take the computations to frequencies higher than about $(2\,\Delta x)^{-1}$.

In addition, observational data are given for a finite range of x, let us say for $-X < x < X$. By a corresponding argument, the Fourier transform need not be calculated for values of s more closely spaced than $\Delta s = (2X)^{-1}$. If there were any significant fine detail in $F(s)$ that required a finer tabulation interval than $(2X)^{-1}$ for its description, then measurements of $f(x)$ would have to be extended beyond $x = X$ to reveal it.

These simple but important facts for the data analyst may be summarized by saying that the function $f(x)$ tabulated at interval Δx over a range $2X$ possesses $2X/\Delta x$ degrees of freedom, and this should be comparable with the number of data computed for the transform. In Chapters 10 and 14 the underlying thought is explored in detail.

A further property of physical data is to contain errors. There is therefore a limit to the precision that is warranted in calculating values of the Fourier transform. This limit is expressed concisely by the power spectrum of the error component (or, what is equivalent, the autocorrelation function of the error compo-

nent). Sometimes only the magnitude of the errors is available, and not their spectrum; and sometimes the magnitude is not known either. Nevertheless, the errors set a limit to the number of physically significant decimal places in a computed value of the transform.

Let $f(x)$ be represented by values at $x = n$, where the integer n ranges from $-X$ to X. Before the Fourier transform of $f(x)$ can be computed, it is necessary to have information about $f(x)$ over the whole infinite range of x. Therefore we call on our physical knowledge and make some assumption about the behavior outside the range of measurement. In this case suppose that $f(x)$ is zero where $|x| > X$. Then the sum

$$\sum_{X}^{X} f(n) e^{-i2\pi sn}$$

will be an approximation to $F(s)$. The real and imaginary parts yield separate sums

$$C = \sum_{-X}^{X} f(n) \cos 2\pi sn \qquad \text{and} \qquad S = \sum_{-X}^{X} f(n) \sin 2\pi sn,$$

and the summations, over $2X + 1$ terms, must be done for each chosen value of s, which can amount to a lot of computing. For this reason it is convenient to possess tables of cosine and sine prepared for suitable values of s; but instead of being limited to one quadrant they should run on and on, according to the size of X, showing negative values as they occur. Because of symmetry, summation from 0 to X suffices for computing as discussed in Chapter 14.

In the mid-sixties the signal processing community was profoundly influenced by Cooley and Tukey (1965). If one divides the data set in two, takes the separate transforms, and combines them, the work can be approximately halved. Further subdivision into even shorter lengths is even better. C. Lanczos had already published on this in 1942, and indeed the idea was being used for numerical Fourier analysis of cometary data by C. F. Gauss in 1805, when he wrote, "Experience will teach the user that this method will greatly lessen the tedium of mechanical calculation." Consequently, after some years the term "Cooley-Tukey Algorithm" was dropped in favor of Fast Fourier Transform. The usual versions of the FFT pretabulate cosines and sines. O. Buneman showed that pretabulation of $\sin\theta$ and $\tan\frac{1}{2}\theta$ was faster (see R. N. Bracewell, 1986), but as processor speeds increase year by year the intrinsic speed of algorithms assumes less importance.

Libraries of subprograms for Fourier analysis and related signal processing and graphical presentation are available in books. See for example W. H. Press et al. (1986) and later volumes for subprograms, on various media, in BASIC, C, FORTRAN, and PASCAL.

High-level software offerings including LINPACK, which is available free from Argonne National Laboratories, MATHEMATICA, FORTRAN 90, and especially MATLAB®, which is student-friendly and which is discussed in Chapter 11, with exercises, in connection with spectral analysis of data sequences, have greatly increased the convenience of access to numerical Fourier analysis.

The single MATLAB statement `F = fft(f);` will operate on a given sequence $\{f\}$ to generate the complex Fourier transform $\{F\}$ virtually instantaneously, whether the number of elements has factors, is prime, or is a power of 2. The inverse transform is performed by `ifft`.

In FORTRAN innumerable subprograms for variations on the fast Fourier algorithm have been published or are available in packages, or "libraries," of general-purpose mathematical subprograms such as IMSL (1980) and NAG (1980) and are invoked by statements of the form `CALL Fourier(f,N,sign)` that replace the N data values of $\{f\}$ by the complex transform values. The forward and inverse transforms respond to the same call according as `sign` is 1 or -1. FORTRAN 90 provides essentially the same facilities as MATLAB.

All the purposes of numerical Fourier transformation may also be achieved by use of the Hartley transform (Chapter 12), which performs both Fourier analysis and synthesis without the need for a separate inverse procedure and does not generate complex output from real input.

Meanwhile, the speed of computers has increased so much that the "slow transform," which directly evaluates sums C and S, may now run fast enough for some purposes. One use is to enable students of spectral analysis to do numerical work before confronting the awkward subtleties of the fast algorithms. Another use in class is to confirm entries in the Pictorial Dictionary and to gain experience with some of the theorems by applying them to those entries. Work depending on algebra or calculus may introduce sign errors, factors of 2 or π, or other slips of hand work, that can be fixed by checking a numerical case faster than by careful repetition of a derivation. Finally, the indispensable blackbox, or canned, routine cannot safely be approached without some understanding of its inscrutable content; comparison against a transparent program is instructive. A common experience is to be visited by a student who has tried to run a canned fast Fourier transform package prematurely and is perplexed when a test run of rectangle-function data does not produce the expected sinc function shape or amplitude.

The code segment following is short, and tolerant of rewriting by the user in language and symbols of preference.

THE SLOW FOURIER TRANSFORM PROGRAM

Let $f(x)$ be given at integer values of x from $-X$ to X, with values of zero being assigned where $f(x)$ is undefined. If integer spacing is unsuitable, rescale the independent variable as wanted. Read the data into an array `f( )`. For computer languages that do not allow the lower array index to be negative, set up a user-defined function to contain the data values.

Since the real and imaginary parts of the transform do not come at integer values of frequency s, an integer array index k, proportional to s and running from 0 to K, is introduced. The frequency s is $k/2K$. Real and imaginary parts $R(k)$ and

$I(k)$ are then computed by summing terms $f(x)^{\cos}_{\sin} 2\pi sx$. Finally, the complex transform is given by $R(k) + iI(k)$ for $k \geq 0$, while for negative k the transform is the complex conjugate $R(k) - iI(k)$. The values chosen for K and X must be stated in advance, the data array **f( )** needs to be filled, and any declarations of variable type, array dimension statements, and other obligatory protocol of the user's language also must be supplied. Here is the program for generating $R(k)$ and $I(k)$.

```
FOR k = 0 TO K          Frequency array index has K + 1 values
    s = k/(2*K)         Frequency variable, maximum is 0.5

    P=2*PI*s
    R(k) = 0            Real part of F(s)
    I(k) = 0            Imaginary part of F(s)
    FOR x = -X TO X     Step in x is unity, limits are ±X
        R(k) = R(k) + f(x)*COS(P*x)
        I(k) = I(k) + f(x)*SIN(P*x)
    NEXT x
NEXT k
```

With many waveforms the choice of origin is unimportant because the wave *shape* does not depend on what instant is chosen as the origin of time. Thus the real and imaginary parts, which are heavily affected by user choice of origin, are not explicitly needed all that often. Instead, R and I are merely steps to getting the power spectrum $R^2 + I^2$, which is an intrinsic property of the waveform itself. The phase spectrum $\arctan(I/R)$ as a function of frequency is also more often needed than R and I; choice of origin does affect phase but only by a simple linear phase shift with frequency.

With data at unity spacing, the shortest period representable is 2; consequently the highest frequency about which information can be obtained from the data is $s_{max} = 0.5$. The longest period that counts is $2X$, the full span of the data; the corresponding lowest frequency is $s_{min} = \frac{1}{2}X$, which also equals Δs.

A choice of $K = X$ will yield $X + 1$ values of R, and X values of I (not counting zero-valued $I(0)$), adding up to exactly the number of independent constants justified by the $2X + 1$ original data values. The spacing $\Delta s_{crit} = \frac{1}{2}K$ over the range $0 \leq s \leq 0.5$ will then produce the number of R and X values that is necessary and sufficient to recover the original data. Call Δs_{crit} the critical spacing. With a choice of $K = 4X$ the spacing Δs is reduced to one-quarter of the critical spacing, and the additional interleaved values computed for $F(s)$, while containing no extra information, will make quite a difference to the appearance of a graph because of the closer packing of points. Conversely, letting k run from 0 to $X/2$ will stretch Δs to double the critical spacing, which is good when debugging a program or when little detail in $F(s)$ is expected.

By letting k run from $5X$ to $7X$ with $K = 12X$ one selects a band around $s = 0.25$ with $\Delta s = 1/24X$; thus a narrow frequency band containing a feature of special interest may be examined at 12 times the critical resolution. This flexibility of the user-created program is not offered by the standard fast algorithms.

Finally, to check transform pairs for gross errors, see that $R(0)$ equals the sum of the data values and that $I(0) = 0$. By not computing R and I for negative

frequency one saves roughly half the running time compared with the standard fast Fourier transform. This time saving also characterizes the Hartley transform.

Example. As input data try

$$\{\tfrac{1}{2}\quad 1\quad 1\quad 1\quad 1\quad 1\quad 1\quad 1\quad 1\quad 1\quad 1\quad 1\quad \tfrac{1}{2}\},$$

representing values of $f(x) = \Pi(x/12)$, with a known transform 12 sinc 12s. The number of data values is 13, with $X = 6$. Choose $K = 2X = 12$, which will yield transform values at half the critical spacing. The program will give

$$R(k) = \{12\quad 7.596\quad 0\quad -2.414\quad 0\quad 1.303\quad 0\quad -.767\quad 0\quad .414\quad 0\quad -.132\quad 0\}$$

and, since $f(x)$ is even, all values of $I(k)$ will be zero. As required, $R(0)$ equals the sum of the data values; check the remaining values against the known transform, noting that $R(1)$ is 0.56% larger than 12 sinc 0.5.

Example. Choose $K = 24$, keeping the data unchanged at $X = 6$. This will give

$$R(k) = \{\mathbf{12}\quad 10.788\quad \mathbf{7.596}\quad 3.555\quad \mathbf{0}\ldots\mathbf{0}\quad -.141\quad \mathbf{-.132}\quad -.046\quad \mathbf{0}\}.$$

The bold values occur at the previous values of frequency s; the interleaved frequencies give a smoother graph. The frequency spacing Δs is now one-quarter of the critical spacing.

Example. Change X to 12 and pad the original data to 25 values by inserting preceding and following zeroes. Keep K at 24. Note that the transform values are not changed by zero padding.

Example. Keep X at 12 but extend the data to 25 values starting and finishing with $\tfrac{1}{2}$ and having 23 ones in between. Keep K at 24. The result is

$$R(k) = 2 \times \{12\quad 7.629\quad 0\quad -2.514\quad 0\quad 1.473\quad 0\quad -1.014\quad 0\quad .749\\ 0\quad -.570\quad 0\quad .439\quad 0\quad -.334\quad 0\}.$$

Note that representation of $\Pi(x)$ by twice as many values reduces the difference between $R(1)$ and 24 sinc 0.5 to 0.14%, a factor of 4 improvement. Verify that the absolute error does not diminish as k increases.

Example. Return to the first example with $K = 12$ and $X = 6$ but change the first line of code to **`FOR k = 0 TO 2 * K`** thereby computing values of $R(k)$ beyond the meaningful limit $s = 0.5$. When k reaches $2K$, s reaches 1.0 and R climbs to the same value that it had at $s = 0$. This exercise shows how the discrepancy between R and 12 sinc 12s arises mainly from the overlapping of a replica $12\,\mathrm{sinc}[12(s-1)]$ and that in general the actual value of R is $\sum_{n=-\infty}^{\infty} 12\,\mathrm{sinc}[12(s-n)]$. Understanding this explains the improvement illustrated in the previous example and shows how to gain improved accuracy by increasing the density of data values.

GENERATION OF TRANSFORMS BY THEOREMS

A wide variety of functions, especially those occurring in theoretical work, can be transformed if some property can be found that permits a simplifying application of a theorem. For example, consider a polygonal function, as in Fig. 7.1. If we perceive that it can be expressed as the convolution of the triangle function and set of impulses, then we can handle the impulses as described above and multiply their transform by the transform of the triangle function. Thus

$$\Lambda(x) * \sum_n A_n \,\delta(x - a_n) \supset \operatorname{sinc}^2 s \sum_n A_n e^{-i2\pi a_n s}.$$

As an example take the trapezoidal pulse $\Lambda(x) * \text{II}(x)$. Evidently

$$\Lambda(x) * \text{II}(x) \supset \operatorname{sinc}^2 s \cos \pi s.$$

In practice the convolution theorem is frequently applicable for the generation of derived transforms, and many examples of the use of the convolution theorem and other theorems will be given in the problems for this chapter.

APPLICATION OF THE DERIVATIVE THEOREM TO SEGMENTED FUNCTIONS

There is a special application of the derivative theorem that has wide use in connection with switching waveforms. Consider a segmentally linear function such

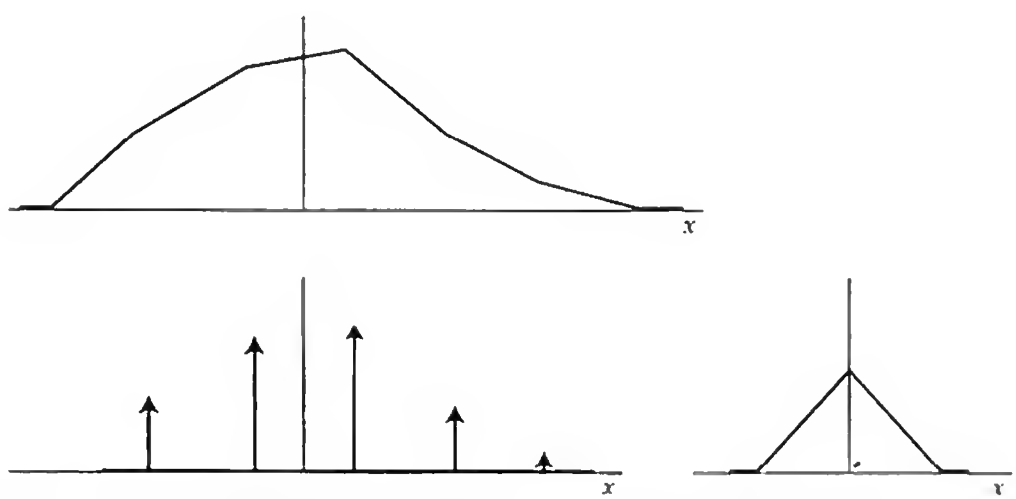


Fig. 7.1 **A polygonal function (above) that can be regarded as the convolution of a set of impulses and a triangle function (below).**

as that of Fig. 7.2*a*. The first derivative, shown in 7.2*b*, contains an impulse. Since the transform of the impulse is known, we remove the impulse and differentiate again (7.2*c*). This time there remains only a set of impulses $\Sigma C_n\,\delta(x - c_n)$. If the first derivative contained, instead of a single impulse as in the example, a set of impulses $\Sigma B_n\,\delta(x - b_n)$, and if the original function $f(x)$ contained impulses $\Sigma A_n\,\delta(x - a_n)$, then evidently the transform $F(s)$ is given by

$$(i2\pi s)^2 F(s) = (i2\pi s)^2 \Sigma A_n e^{-i2\pi a_n s} + i2\pi s \Sigma B_n e^{-i2\pi b_n s} + \Sigma C_n e^{-i2\pi c_n s}.$$

Clearly this technique extends immediately to functions composed of segments of polynomials, in which case further continued differentiation is required. As a simple example let us consider the parabolic pulse $(1 - x^2)\Pi(x/2)$. Here

$$(1 - x^2)\Pi\left(\frac{x}{2}\right) \supset F(s).$$

Differentiating twice, we have

$$-2x\Pi\left(\frac{x}{2}\right) \supset i2\pi s F(s)$$

$$2\delta(x + 1) - 2\Pi\left(\frac{x}{2}\right) + 2\delta(x - 1) \supset (i2\pi s)^2 F(s).$$

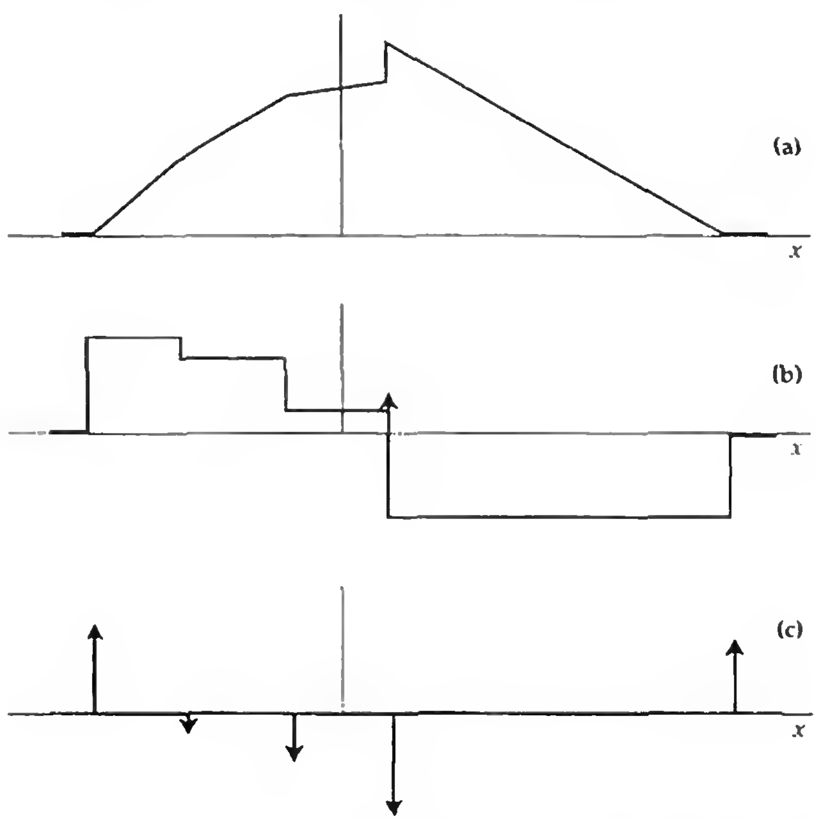


Fig. 7.2 Technique of reduction to impulses by continued differentiation.

Now the left-hand side has a known transform $4 \cos 2\pi s - 4 \operatorname{sinc} 2s$. Hence the desired transform $F(s)$ is given by

$$F(s) = \frac{4 \cos 2\pi s - 4 \operatorname{sinc} 2s}{(i2\pi s)^2} + K_1\delta(s) + K_2\delta'(s),$$

where K_1 and K_2 are integration constants arising from the fact that a constant K_1 may be added to the original parabolic pulse without changing its first derivative, and similarly for the second derivative. Integration of the given pulse shows that no additive constant or linear ramp is present, so in this case $K_1 = K_2 = 0$ and

$$F(s) = -\frac{\cos 2\pi s}{\pi^2 s^2} + \frac{\sin 2\pi s}{2\pi^3 s^2}.$$

MEASUREMENT OF SPECTRA

While spectra can be determined mathematically either from waveforms expressed algebraically or from discrete numerical data, the historical sequence of events begins with Isaac Newton as explained in his *Opticks* of 1704. Spectra are still produced by prisms and other optical devices, especially diffraction gratings and Fabry-Perot interferometers.

The following sections deal with two special cases of the determination of spectra from signals presented in physical form.

Radiofrequency spectral analysis. Radio communication, radar, television, and laboratory instruments all make use of spectrum analyzers to display signal strength as a function of frequency. A radio receiver is being used as a primitive spectrum analyzer when the tuning knob is being turned; a commercial device can be imagined that mechanically varies the capacitance of a tuned circuit and displays the response on a cathode ray tube whose spot is moved horizontally in synchronism with the variation of the tuning capacitor. Such an instrument, even when voltage-controlled capacitors have eliminated moving parts, has only a limited bandwidth. With the passage of time impressive advances have been made in digitizing oscilloscopes, so that the frequency band from d.c. to millimeter wavelengths can be covered with virtually level sensitivity. Such an instrument samples the waveform under study with conventional electric circuitry, performs fast spectral analysis by computer, and displays or stores the spectrum or any other desired product of digital signal processing. By 1990 fully programmable instrumentation based on the Hartley transform was commercially available with picosecond resolution.

An alternative procedure is to employ a bank of fixed-frequency filters, implemented by computer to be sure, as exemplified by the SETI Institute in searching for radio signals from nonsolar planets over a 20 MHz band with resolution as fine as 1 Hz.

Optical Fourier transform spectroscopy. In the optical wavelength range spectrum analysis has been carried out by massively parallel analogue devices, namely prisms and diffraction gratings. But there is another way, not in the line of evolution from electric circuit practice, and it has become of particular importance for detecting and identifying molecules by their infrared spectra. If one could determine the temporal autocorrelation of an electromagnetic signal $s(t)$ then by Fourier transformation one would have the power spectrum. The technique for doing this is to split the signal to be analyzed into two beams with a view to forming the product $s(t)s(t + \tau)$, where τ is a variable time delay. Beam splitting can be accomplished with high efficiency by a mylar film oriented at 45°. One beam continues straight through with half the incident power; the other beam reflects at a right angle. Before the two beams are allowed to recombine and pass to a detector, a relative delay τ is inserted in one beam. The beam to be delayed is diverted towards a plane mirror and after reflection, returns to its original path. If now the mirror is moved away then an increasing time delay τ is introduced. With a maximum delay τ_{max} a spectral resolution bandwidth of $\frac{1}{2}\tau_{max}$ is achieved. The total band covered extends to a high frequency $\frac{1}{2}\Delta\tau$, where $\Delta\tau$ is the time for which the moving mirror dwells at each value τ. If the analysis has to be done fast, as when dynamic chemical reactions are to be studied, the mirror can be moved continuously, even at explosive speeds and a loud bang. In this case $\Delta\tau$ is the change in τ in one integration time of the final detected output signal. At the highest mechanical velocities and repetition rates required for transient analysis, linear motion cannot be counted upon. Instead, the mirror position is followed by counting optical interference fringes, or alternatively moiré fringes, a count that is used to synchronize digitization at the detector output with the position of the moving mirror.

It would suffice to work with positive delays only, since the autocorrelation function is even, and the time-averaged $s(t)s(t + \tau)$ equals that of $s(t - \tau)s(t)$. However, if τ is varied from positive to negative then the origin of the autocorrelogram is apparent in the presence of small maladjustments. This is a practical point, but also supports a different function of the instrument.

If instead of determining the spectrum of the incident radiation one uses a known spectrum, for example from a black-body radiator, then a transparent sample may be placed in one beam and a cross-correlogram will be recorded. This is not an even function; Fourier transformation will now deliver a complex result from which properties of the sample, such as permittivity and conductivity, or refractive index and absorption coefficient, can be found over the full spectral range. This is a richer result than a simple absorption spectrum. Given an opaque specimen, one can use it as the moving mirror, determine the complex reflection coefficient, and thence electromagnetic properties including conductivity, permittivity, and penetration depth (Bell, 1972).

Although instruments for spectral analysis have originated and been developed for use in applications outside the range of electrical engineering, manufacture has been in the hands of the engineers who have brought us instruments for medical diagnosis, medical treatment, clocks, and navigational devices, illustrating what Lord Kelvin said about Fourier's Theorem.

BIBLIOGRAPHY

Bell, R. J.: "Introduction to Fourier Transform Spectroscopy," Academic Press, New York, 1972.

Bracewell, R. N.: "The Hartley Transform," Oxford University Press, New York, 1986.

Cooley, J. W., and J. W. Tukey: "An Algorithm for the Machine Calculation of Complex Series," *Math. of Comput.*, vol. 19, pp. 297–301, 1965.

IMSL: "IMSL Library Reference Manual," IMSL Inc., Houston, Texas, 1980.

NAG: "NAG Fortran Library Manual Mark 8," NAG Central Office, Oxford, England, 1980.

Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling: "Numerical Recipes, the Art of Scientific Programming," Cambridge University Press, Cambridge, U.K., 1990. See also successor volumes devoted to programs and examples in BASIC, C, FORTRAN 77, FORTRAN 90, and PASCAL.

PROBLEMS

1. **Checking the second moment.** See whether the exact relation $\int_{-\infty}^{\infty} x^2 f(x)\,dx = -F''(0)/4\pi^2$ can be made the basis for a numerical check of $R(k)$, the computed real part of the transform of $f(x)$ (as sampled at unit intervals of x over the range $-X$ to X. The easily computed second moment $\sum_{-X}^{X} x^2 f(x)$ should agree with what property of $R(k)$? ▷

2. **Transform of $\cos \pi x^2$.** Evaluate $\int_{-\infty}^{\infty} \cos \pi x^2 \cos 2\pi s x\,dx$ to show that the Fourier transform of $\cos \pi x^2$ is $F(s) = 2^{-1/2}(\cos \pi s^2 + \sin \pi s^2)$. Before attempting the integration think how you could apply a numerical check in case a gross error has been made by omission of a factor such as $\sqrt{2}$ or $\sqrt{\pi}$. ▷

3. **Transform of $\exp i\pi x^2$.** Show that
$$e^{i\pi x^2} \supset \sqrt{i}\,e^{-i\pi s^2}. \;▷$$

4. **A more symmetrical transform.** Show that
$$e^{i\pi(x^2-\frac{1}{8})} \supset e^{-i\pi(s^2-\frac{1}{8})}.$$

5. **Unsymmetrical function.** A function of x that is nonzero only for $-16 < x < 16$, and is neither even nor odd, is $\Lambda(x/16) + \Lambda(x/8 + 1)$. (*a*) Confirm that its exact Fourier transform is $F(s) = 16 \operatorname{sinc}^2 16s + 8 \operatorname{sinc}^2 8s \exp(i2\pi 8s)$. (*b*) Tabulate the 33 values of $f(x)$ for $x = -16, -15, \ldots 16$ and use them in the slow F.T. program to compute the real and imaginary parts of the transform for $0 \le s \le 0.5$ with $\Delta s = 1/64$, which is half the critical spacing that the FFT would generate. (*c*) Comment on the agreement of the results with the exact real and imaginary parts of $F(s)$. ▷

6. **Unsymmetrical function.** Values of a certain $f(i)$ for $i = -4$ to 4 are $\{1\ \ 2\ \ 3\ \ 4\ \ 5\ \ 6\ \ 7\ \ 8\ \ 9\}$. Use the slow Fourier Transform for values of s running from 0 to 0.5 at intervals $\Delta s = 1/18$ and compare the results with the transform of $g(x) = x\Pi[(x-5)/9]$. ▷

7. **Central limit theorem?** Does the product $(1 - s^2)(1 - s^2/4)(1 - s^2/9)\ldots(1 - s^2/N^2)$ approach Gaussian form for $s << 1$ as N becomes large? ▷

8. **Weekly summing.** The occurrence of certain events is dependent on, or correlates with, the day of the week, examples ranging over happenings as diverse as emergency-room admissions, atmospheric clarity, and noise pollution. Summing the number of events as a function of the day of the week provides a statistical tool for investigation and possible mitigation. Given a data set g_j for $j = 0$ to 6, accumulated in this way over a whole number of weeks, an investigator considers Fourier analysis. Since there are exactly seven data values, no more than seven independent numbers F_k ($k = 0$ to 6) should be generated by such a change in mode of presentation. In the case of a data set $g_j = \{5 \quad 4 \quad 9 \quad 8 \quad 7 \quad 6 \quad 10\}$ from Sunday to Saturday, what are the seven numbers F_k? ▷

9. **Sine transform.** Find the sine transform $F_s(s)$ for $f(x) = \Lambda(x/a - 1)$.

10. **Cosine transform.** Find the cosine transform $F_c(s)$ for $f(x) = \Lambda(x/a - 1)$.

8

The Two Domains

We may think of functions and their transforms as occupying two domains, sometimes referred to as the upper and the lower, as if functions circulated at ground level and their transforms in the underworld (Doetsch, 1943). There is a certain convenience in picturing a function as accompanied by a counterpart in another domain, a kind of shadow which is associated uniquely with the function through the Fourier transformation, and which changes as the function changes. In the illustrations given here the uniform practice is to keep the functions on the left and the transforms on the right.

The theorems given earlier can be regarded as a list of pairs of corresponding simultaneous operations, one in the function domain and the other in the transform domain. For example, compression of the abscissas in the function domain means expansion of the abscissas plus contraction of the ordinates in the transform domain; translation in the function domain involves a certain kind of twisting in the transform domain; and convolution in the function domain involves multiplication in the transform domain.

By applying these theorems to everyday problems we are able at will to cross from one domain to the other and to carry out required operations in whichever domain is more advantageous. We may find on reaching the conclusion to a line of argument that we are in the wrong domain; our conclusion is a statement about the transform of the function we are interested in.

We therefore now consider pairs of corresponding properties; the area under a function and the central ordinate of its transform are such a pair. Suppose that we are interested in the area under a function, but we have only its transform; the desired quantity proves to be equal to the central ordinate of the transform. It is not necessary for us to invert the Fourier transformation in full and to integrate the function so determined.

It very often happens that particular features of a function are all we need to know. The corresponding pairs of properties assembled here then assume an im-

portant role in enabling us to sidestep details of transformations and go to a direct answer.

DEFINITE INTEGRAL

The definite integral of a function from $-\infty$ to ∞ is equal to the value of its transform at the origin; that is,

$$\int_{-\infty}^{\infty} f(x)\,dx = F(0).$$

Derivation:

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{\infty} f(x)e^{-i2\pi xs}\,dx\bigg|_{s=0} = F(0).$$

We often refer to this integral as the area under a function; thus in Fig. 8.1 it is the area shown shaded. The term "central ordinate" is used for the value at the origin of abscissas and this value is indicated in the figure by a heavy line.

It follows from the symmetry of the Fourier transform that the area under the transform is equal to $f(0)$. From this simple correspondence of properties many very interesting conclusions can be drawn. For instance, any operation on $f(x)$ which leaves its area unchanged—say a translation along the x axis—leaves the central ordinate of the transform unchanged. In the case of translation, this conclusion is verifiable from the shift theorem, for

$$\overline{f(x-a)}|_0 = e^{-i2\pi as}F(s)|_0 = F(0).$$

A more abstruse example is furnished by $f(x-a)H(x-a) + f(x+a)H(-x-a)$, where a given $f(x)$ is separated at $x = 0$ into two equal parts, which are then pushed apart by an amount $2a$. Since the area has not been changed, the central ordinate of the transform is the same. On the other hand, since the central ordinate of the function is now zero, we can say that the area of the transform must now be zero. These two special facts about the transform are a good example of the powerful reasoning based on corresponding properties.

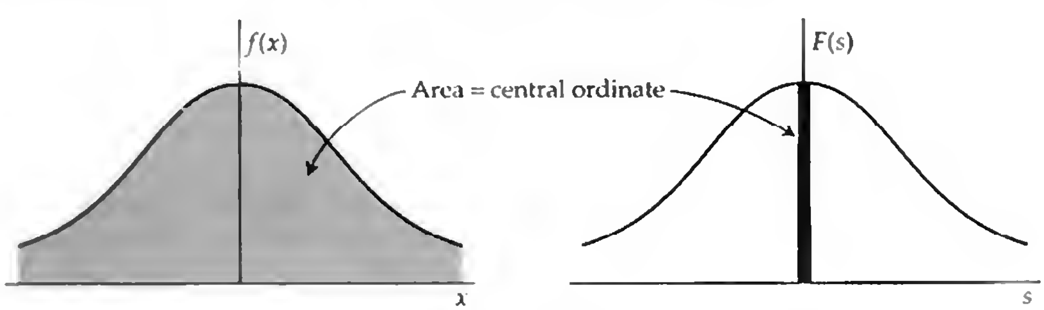


Fig. 8.1 The area under a function is equal to the central ordinate of its transform.

TABLE 8.1
Some functions and their areas

$f(x)$	$F(0)$	$f(x)$	$F(0)$
$\Pi(x)$	1	$\cos x$	0
$\operatorname{sinc} x$	1	$\text{Ⅱ}(x)$	0
$\Lambda(x)$	1	$\sin x$	0
$\operatorname{sinc}^2 x$	1	$\text{III}(x)$	∞
$e^{-\pi x^2}$	1	$\delta(x)$	1
$J_0(x)$	1	e^{ix}	0
$\dfrac{J_1(x)}{x}$	1	$H(x)$	∞
$\text{ıı}(x)$	1	$\operatorname{sgn} x$	0

Exercise. If $f(x)$ were a rectangle function, then the foregoing operation of splitting and separating the parts would result in a rectangular pulse pair. Obtain the full transform by use of the modulation theorem, and verify that the infinite integral of the spectrum of a pulse pair is indeed zero.

Exercise. What distinctive character can you assign to the spectrum of a signal in Morse code?

In Table 8.1 many special cases are collected, the right-hand column showing the area under the function to the left.

Exercise. The functions $\cos x$ and $\sin x$ have infinite integrals which are oscillatory, yet the central ordinates of their transforms have a perfectly definite value of zero. However, the transforms are transforms in the limit. The definite-integral/central-ordinate relation, when applied to transforms in the limit, must thus imply some rule for integrating (co)sinusoids. Investigate this matter. Hint: Consider transformable functions such as $\cos x\, \Pi(x/M)$ and $\cos x\, \Pi[(x - \frac{1}{4}\pi)/M]$, with the aid of the convolution and shift theorems, in the limit as $M \to \infty$.

Exercise. Consider what would be appropriate entries in the table for $H(x)$ and $\operatorname{sgn} x$.

THE FIRST MOMENT

By analogy with mass distribution along a line, the first moment of $f(x)$ about the origin is defined as

$$\int_{-\infty}^{\infty} xf(x)\,dx.$$

The following theorem connects the first moment of a function with the central slope of its transform.

If $f(x)$ has the Fourier transform $F(s)$ then the first moment of $f(x)$ is equal to $-(2\pi i)^{-1}$ times the slope of $F(s)$ at $s = 0$; that is,

$$\int_{-\infty}^{\infty} xf(x)\,dx = \frac{F'(0)}{-2\pi i}.$$

Derivation: By the derivative theorem,

$$\int_{-\infty}^{\infty} (-2\pi ix)f(x)e^{-i2\pi xs}\,dx = F'(s).$$

Therefore $$-2\pi i\int_{-\infty}^{\infty} xf(x)\,dx = F'(0).$$

In Fig. 8.2 the impulse pair, $-Ⅱ(x)$ is shown with its transform $-i\sin\pi s$. Each impulse has a positive moment $\frac{1}{4}$, hence the total moment is $\frac{1}{2}$; thus

$$-\int_{-\infty}^{\infty} xⅡ(x)\,dx = -\int_{-\infty}^{\infty} x\tfrac{1}{2}\delta(x+\tfrac{1}{2})\,dx + \int_{-\infty}^{\infty} x\tfrac{1}{2}\delta(x-\tfrac{1}{2})\,dx = \tfrac{1}{2}.$$

Hence $$F'(0) = (-2\pi i)\tfrac{1}{2} = -i\pi.$$

In Fig. 8.3 a truncated exponential and its transform are shown. The derivative of the real part of the transform is zero at the origin, so the "slope at the origin" is given by the derivative $-i\tan\theta$ of the imaginary part alone. The first moment of the exponential distribution is thus given by $(\tan\theta)/2\pi$, a positive real quantity.

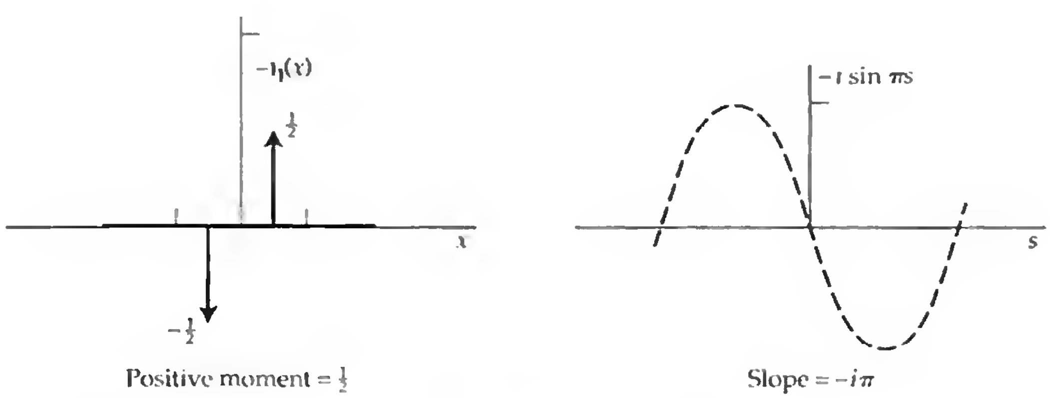


Fig. 8.2 **The moment of a function is proportional to the central slope of the transform.**

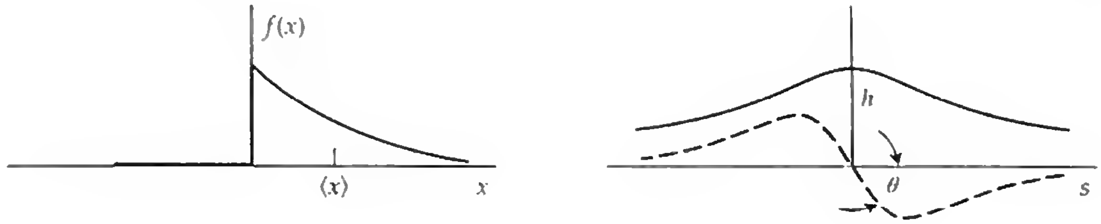


Fig. 8.3 **The abscissa of the center of gravity is equal to (tan θ)/$2\pi h$.**

As a special case consider a function whose first moment is zero. Its transform has zero slope at its origin. Conversely, *if a function has zero slope at the origin, its transform has zero first moment.* Furthermore, if a function has zero slope at the origin and is not zero there, then its transform has its center of gravity at the origin.

Exercise. Discuss the first moment and center of gravity of the transform of $\Pi(x - a)$ as a goes from zero to infinity.

Exercise. The function $(x - z)\Pi(x - z)$ is zero and has zero slope at the origin, and it has zero area. What can be said about its transform?

Exercise. Under what condition on $f(x)$ will the derivative of $F(s)$ be purely imaginary at $s = 0$?

CENTROID

By the centroid of $f(x)$ we mean the point with abscissa $\langle x \rangle$ such that the area of the function times $\langle x \rangle$ is equal to the first moment. Thus

$$\langle x \rangle = \frac{\int_{-\infty}^{\infty} xf(x)\,dx}{\int_{-\infty}^{\infty} f(x)\,dx}.$$

Roughly speaking, $\langle x \rangle$ tells us where a function $f(x)$ is mainly concentrated or, when a signal pulse is described as a function of time, $\langle t \rangle$ bears on the epoch of the signal. In statics, $\langle x \rangle$ is the abscissa of the center of gravity of a rod whose mass density is $f(x)$. Late we shall find that the centroid of $[f(x)]^2$ is a more suitable measure of location or epoch in cases where $f(x)$ can go negative. In the worst case the area of $f(x)$ is zero and $\langle x \rangle$ becomes infinite.

Since $\langle x \rangle$ is the ratio of the first moment to the area of $f(x)$, its connection with $F(s)$ follows immediately from the preceding sections:

$$\langle x \rangle = -\frac{F'(0)}{2\pi i F(0)}.$$

In other words, *the abscissa of the centroid of a function is given by the central slope over the central ordinate of its transform,* the constant of proportionality being $-(2\pi i)^{-1}$.

If the example of Fig. 8.2, chosen to illustrate first moments, the area under $f(x)$ is zero, and so the idea of center of gravity fails. To give an analogy from mechanics, if $h(x)$ described a distribution of forces applied to a lever, then no balance point could be found to which an equilibrating force could be applied.

In Fig. 8.3, however, the center of gravity of $f(x)$ falls at a finite positive value of x. Since $F(s)$ is hermitian, its derivative at the origin is purely imaginary, say, $i \tan \theta$; the central ordinate h is given by the value of the real part of $F(s)$ alone; thus

$$\langle x \rangle = \frac{\tan \theta}{2\pi h}.$$

MOMENT OF INERTIA (SECOND MOMENT)

The moment of inertia or second moment of a function $f(x)$ about the origin is defined as

$$\int_{-\infty}^{\infty} x^2 f(x)\, dx.$$

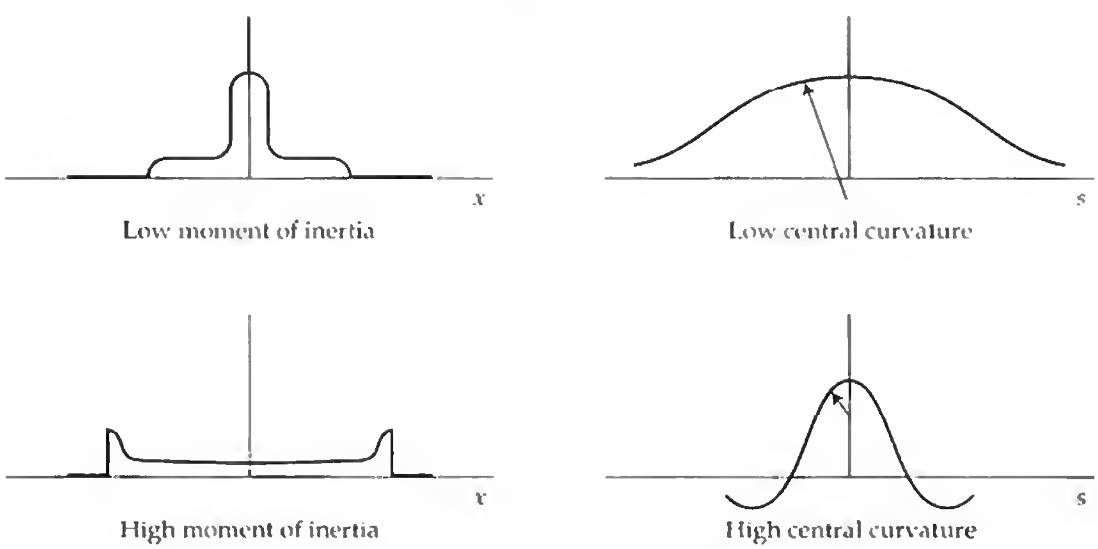


Fig. 8.4 **The moment of inertia is proportional to the central curvature of the transform.**

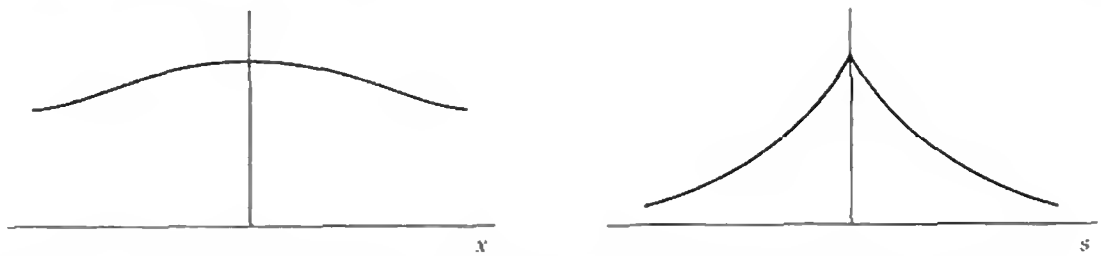


Fig. 8.5 **A function with an infinite second moment has a transform with a discontinuity in slope at the origin.**

One finds that the higher the second moment of a function the more curved its transform at the origin (Fig. 8.4). Because of the factor x^2, it is clear that the second moment of a function is sensitive to the behavior of $f(x)$ at large values of x; thus what happens at *large* x is reflected in the behavior of the transform at the *origin* of s. This remark also applies to the theorem on the first moment.

The moment of inertia of $f(x)$ is equal to $(4\pi^2)^{-1}$ times the downward curvature of its transform at the origin; that is,

$$\int_{-\infty}^{\infty} x^2 f(x)\, dx = -\frac{1}{4\pi^2} F''(0).$$

Derivation: By the derivative theorem,

$$\int_{-\infty}^{\infty} (-i2\pi x)^2 f(x) e^{-i2\pi xs}\, dx = F''(s).$$

Therefore
$$\int_{-\infty}^{\infty} x^2 f(x)\, dx = -\frac{F''(0)}{4\pi^2}.$$

A special situation of frequent occurrence is shown in Fig. 8.5, where a function whose second moment is infinite has a transform whose central curvature is infinite, that is, the slope jumps abruptly at the origin.

MOMENTS

The nth moment of $f(x)$ is equal to $(-2\pi i)^{-n}$ times the nth differential coefficient of $F(s)$ at the origin; that is,

$$\int_{-\infty}^{\infty} x^n f(x)\, dx = \frac{F^{(n)}(0)}{(-2\pi i)^n},$$

provided that the moment exists.

If $f(x)$ behaves as $|x|^{-m}$ for large $|x|$, where m is positive, then only the first $[m]$ moments will exist, and only the first $[m]$ differential coefficients of $F(s)$ at $s = 0$ will exist; there will be a discontinuity of the $[m + 1]$th order at $s = 0$. (The function $[m]$ is the greatest integer less than m.)

For example, if m is a positive integer M, and $f(x)$ dies away as x^{-M}, then the Mth moment is infinite and the Mth derivative of $F(s)$ is impulsive. Conversely, if the Mth derivative of $f(x)$ is impulsive, then its transform $F(s)$ dies away as s^{-M}

MEAN-SQUARE ABSCISSA

The mean-square abscissa, which will be represented by $\langle x^2 \rangle$ is the mean value of x^2, the mean being weighted according to the distribution $f(x)$. Thus

$$\langle x^2 \rangle \int_{-\infty}^{\infty} f(x)\, dx = \int_{-\infty}^{\infty} x^2 f(x)\, dx,$$

and hence by previous results

$$\langle x^2 \rangle = -\frac{F''(0)}{4\pi^2 F(0)}.$$

In dynamics, the square of the radius of gyration of a rod whose mass density is $f(x)$ is the same as $\langle x^2 \rangle$. In statistics, $\langle x^2 \rangle$ for a frequency distribution function $f(x)$ is an important concept [the same as the variance of the function $f(x)$] and it is useful to know the relationship it bears to the Fourier transform of $f(x)$ (the "characteristic function"). The mean-square abscissa possesses a very interesting additive property under convolution.

The mean-square abscissa for $f * g$ is equal to the sum of the mean-square abscissas for f and g, provided that f (or g) has its centroid at its origin.

We expect the convolution $f * g$ to be wider, in some sense, than either of its components; this theorem is one quantitative expression of the smoothing-out or diffusing effect of convolution.

Derivation:

$$\begin{aligned}\int_{-\infty}^{\infty} x^2(f * g)\, dx &= \overline{x^2(f * g)}\Big|_0 = -(4\pi^2)^{-1}\overline{(f * g)''}\Big|_0 \\ &= -(4\pi^2)^{-1}(FG)''\Big|_0 \\ &= -(4\pi^2)^{-1}[F''G + 2F'G' + FG'']_0\end{aligned}$$

$$\int_{-\infty}^{\infty} (f * g)\, dx = FG\Big|_0$$

$$\langle x^2\rangle_{f*g} = \frac{\int_{-\infty}^{\infty} x^2(f*g)\,dx}{\int_{-\infty}^{\infty}(f*g)\,dx} = -(4\pi^2)^{-1}\left[\frac{F''}{F} + 2\frac{F'G'}{FG} + \frac{G''}{G}\right]_0$$

$$= \langle x^2\rangle_f + \langle x^2\rangle_g + 2\frac{F'(0)}{2\pi i F(0)}\frac{G'(0)}{2\pi i G(0)}.$$

The final term on the right-hand side, which is twice the product of the abscissas of the centroids of f and g, will vanish if either f or g has its centroid at its origin.

By taking $g(x) = \delta(x - a)$, for which $\langle x^2\rangle_g = a^2$, we have the theorem which is familiar in relation to the moment of inertia of a mass about an axis not through its center of gravity. For if $f(x)$ has its origin at its centroid,

$$\langle x^2\rangle_{f(x-a)} = \langle x^2\rangle_f + a^2.$$

Other ways of weighting the means of x^2 may be considered. For example, if $f(x)$ has negative-going parts, they will reduce the variance, but there may be interest in measuring the spread of $f(x)$ irrespective of its sign. Two quantities which are met in these circumstances are

$$\langle x^2\rangle_{|f|} \quad \text{and} \quad \langle x^2\rangle_{|f|^2}.$$

RADIUS OF GYRATION

This term stands for the root-mean-square value of x—that is, the square root of the mean-square abscissa—and its convenience is largely associated with having the same dimensions as x. Properties such as the relation to the central curvature of the transform, the additive property, and the radius of gyration about a displaced axis, are usually simpler when expressed in terms of the mean-square abscissa rather than in terms of its square root.

VARIANCE

The variance σ^2 is the mean-square deviation referred to the centroid; that is,

$$\sigma^2 = \langle(x - \langle x\rangle)^2\rangle = \frac{\int_{-\infty}^{\infty}(x - \langle x\rangle)^2 f(x)\,dx}{\int_{-\infty}^{\infty} f(x)\,dx}$$

$$= -\frac{F''(0)}{4\pi^2 F(0)} + \frac{[F'(0)]^2}{4\pi^2[F(0)]^2},$$

and the statements about $\langle x^2\rangle$ apply also to variance, since a choice of the origin of abscissas which makes $\langle x\rangle$ zero will mean that $\langle x^2\rangle$ is the same as the variance.

The variance of $f * g$ is equal to the sum of the variances of f and g.

The proof follows from the earlier result for mean-square abscissas.

SMOOTHNESS AND COMPACTNESS

The smoother a function is, as measured by the number of continuous derivatives it possesses, the more compact is its transform; that is, the faster it dies away with increasing s.

If a function and its first $n - 1$ derivatives are continuous, its transform dies away at least as rapidly as $|s|^{-(n+1)}$ for large s; that is,

$$\lim_{|s|\to\infty} |s|^n F(s) = 0.$$

In the customary symbology of Landau, $F(s) = O(|s|^{-n})$. In the course of the proof of this theorem the Fourier transforms of the first $n - 1$ derivatives appear, and it is therefore necessary to mention, in a rigorous statement of the theorem, that the first n derivatives should possess absolutely convergent integrals.

In Fig. 8.6 several of the common cases are illustrated. The rather familiar fact that $\text{sinc}^2 x$ dies away much more rapidly than sinc x is seen to be associated with the smoother behavior of its transform. We can say that functions possessing discontinuities in slope ("corners"), and whose second derivatives are therefore impulsive, will have transforms that die away as $|s|^{-2}$. In general one can say that if the kth derivative becomes impulsive, then the transform behaves as $|s|^{-k}$ at infinity. This simplifies statements, such as the one given above, which contain $n - 1$, n, and $n + 1$ in the same sentence.

It is interesting to note that the function $\exp(-\pi x^2)$ is continuous and has all continuous derivatives, and is therefore in the present sense as smooth as possible. Likewise its transform $\exp(-\pi s^2)$ is as compact as possible, since it dies away faster than s^{-n} for all n. This property is not unique to $\exp(-\pi x^2)$, being shared with many other pairs of transforms (including, for example, the self-reciprocal pair sech πx, sech πs).

To prove the above statement, let $f(x)$ have an absolutely integrable derivative [$f(x)$ is of course itself absolutely integrable and differentiable]. Since $f'(x)$ is absolutely integrable,

$$g(x) = \int_{-\infty}^{x} f'(\xi)\, d\xi$$

exists, and since $g'(x) = f'(x)$,

$$f(x) = g(x) + c.$$

As $x \to -\infty$, $g(x) \to 0$. Now if c were not zero, $f(x)$ could not be absolutely integrable. Hence $g(x) = f(x)$, and thus as $x \to -\infty$, we also have $f(x) \to 0$.

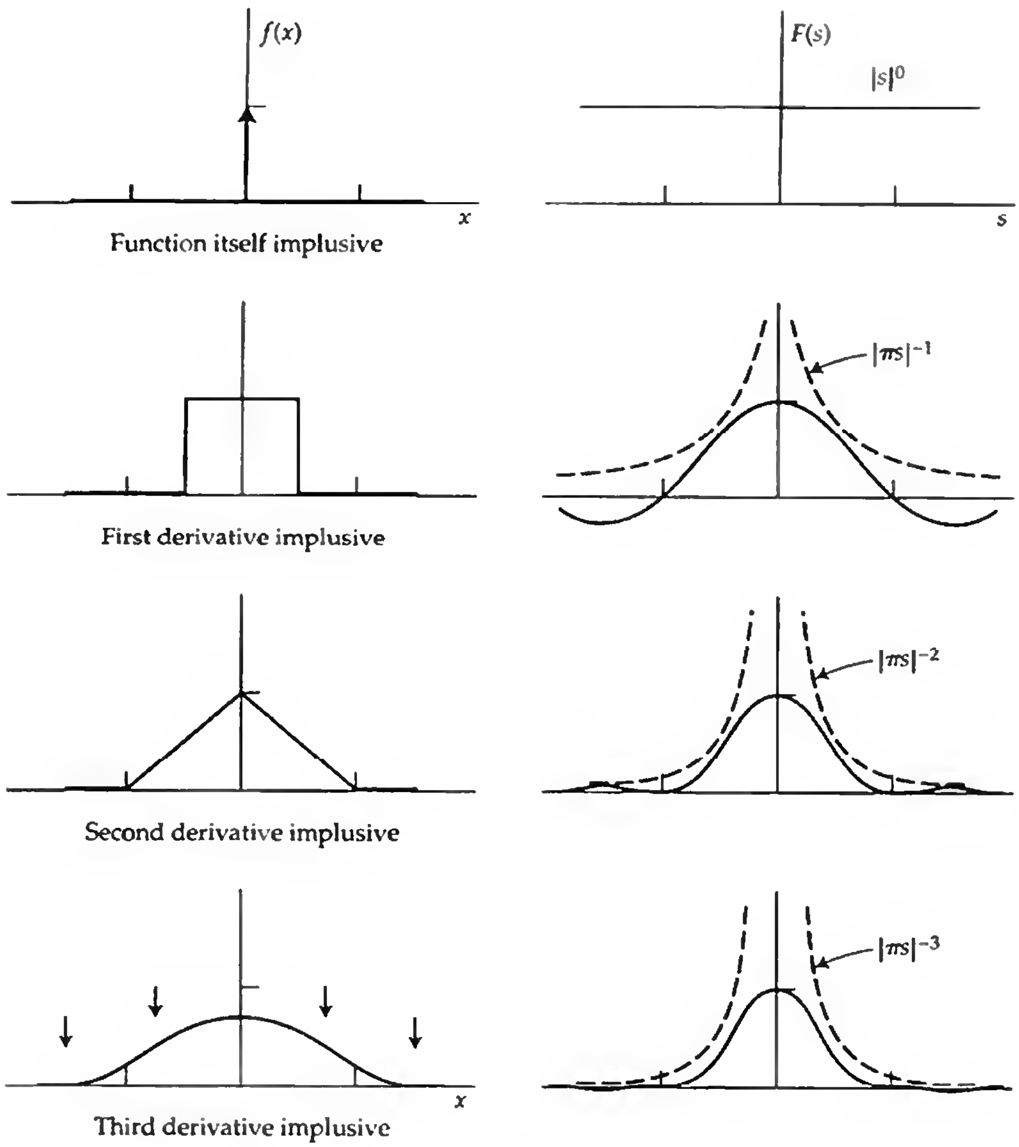


Fig. 8.6 **If the kth derivative of a function becomes impulsive, its transform behaves as $|s|^{-k}$ at infinity.**

Now put

$$\int_{-\infty}^{\infty} f'(\xi)\, d\xi = c_1;$$

then

$$f(x) = \int_{-\infty}^{x} f'(\xi)\, d\xi = c_1 - \int_{x}^{\infty} f'(\xi)\, d\xi.$$

So again $c_1 = 0$ and thus $f(x) \to 0$ for $x \to \infty$. Now

$$F(s) = \int_{-\infty}^{\infty} f(x) e^{-i2\pi sx}\, dx$$

$$= \frac{\int_{-\infty}^{\infty} f'(x)e^{-i2\pi sx}\,dx}{i2\pi s},$$

on integrating by parts and using the fact that $f(x) \to 0$ as $x \to \pm\infty$.

Then

$$sF(s) \to 0 \qquad \text{as} \qquad s \to \pm\infty$$

$$sF(s) = o(|s|^{-1})$$

or

$$F(s) = o(|s|^{-2}).$$

If then, continuing the argument, $f(x)$ possesses n absolutely integrable derivatives, we have $F(s) = o(|s|^{-n})$.

SMOOTHNESS UNDER CONVOLUTION

We know from numerical experience that convolution produces broadening and smoothing effects. This is also expected on physical grounds in situations such as scanning the sound track of a film, where convolution describes the physical phenomenon. One way of giving a precise definition to the general idea of broadening has already been established; thus it is known that the standard deviation σ_f, which is a measure of spread of a function $f(x)$, is increased under convolution with $g(x)$ according to the result

$$\sigma_{f*g}^2 = \sigma_f^2 + \sigma_g^2.$$

There are other ways of measuring spread, and smoothness too could be measured in different ways. If an automobile were driven along a road whose height varied as Π^{*2}, or $\Lambda(x)$, the ride would not be smooth; but it would be smoother than driving along a road described by $\Pi(x)$, which has cliffs. A road surface Π^{*3}, composed of two straight and three parabolic segments, free from abrupt changes in height or slope, would be smoother. It would not be perfectly smooth, for at each of the four points of articulation there would be an abrupt change in the centrifugal force, associated with the abrupt change in curvature. In these cases the smoothness increases with the number of derivatives possessed by the function.

Let us say that if the kth derivative of a function is impulsive, then it has smoothness of order k. Now let $f(x)$ have smoothness of order m, and $g(x)$ smoothness of order n. When they are convolved together, is the convolution indeed smoother, as we would wish, under our quantitative definition?

Since

$$F(s) \sim s^{-m}$$

and

$$G(s) \sim s^{-n}$$

it follows that

$$F(s)G(s) \sim s^{-(m+n)}.$$

But $F(s)G(s)$ is the Fourier transform of $f * g$, hence the $(m + n)$th derivative of $f * g$ is impulsive, and it therefore has smoothness of order $m + n$. It is thus smoother than the functions entering into the convolution; in fact the smoothness proves to be additive under convolution just as σ^2 is.

The word "smoothing" is sometimes used as a synonym for convolution. "Inverse smoothing" and "sharpening" are terms used to mean solving the convolution integral equation. It is rather interesting from the standpoint of this section that "sharpening" procedures make use of further smoothing. Consequently the quantitative measure here given to smoothness does not always coincide with qualitative conceptions of smoothness. As a particular example, consider $\Lambda(x) * \tau^{-1} \exp(-\pi x^2/\tau^2)$, where τ is very small. This function is now infinitely smooth but is a close approximation to the rough $\Lambda(x)$. It differs appreciably only at the corners, which are microscopically rounded off to prevent impulsive derivatives. The new road profile might, however, be just as rough to ride over as the old.

ASYMPTOTIC BEHAVIOR

It is well known qualitatively that sharp or abrupt features in a waveform betoken the presence of high-frequency components. Now we can go further and say quantitatively how the spectrum holds up as frequency increases.

We have seen that if a function is continuous but has a discontinuous derivative—that is, it has no jumps but has corners—then its transform dies away as s^{-2} and its power spectrum as s^{-4}. To describe such behavior in electric-circuit language, we would refer to it as attenuation at 12 decibels per octave or 40 decibels per decade.

On a log-log plot, a function varying as s^{-2} would become a straight line of slope -2. Logarithmic units such as the octave or decade, the decibel, the neper, and the magnitude (in astronomy) permit slope, which is dimensionless in itself, to be stated in expressive dimensionless units.

Figure 8.7 shows linear and log-log versions of two familiar functions. (Where the function is negative, the logarithm of the absolute value is shown.) Certain aspects of the functions, especially the nature of their asymptotic behavior and their closeness of approach thereto, are more clearly evidenced by the logarithmic than by the linear plot. For this and other reasons, logarithmic plots are often used to display electrical-filter characteristics and antenna patterns.

The question arises why only slopes of 0, 3, 6, . . . , decibels per octave were encountered when discontinuities of various kinds were discussed in the preceding section. We can study this by considering some transform which dies away at 4.5 decibels per octave, that is, as $s^{-3/4}$, and inquiring what sort of discontinuity the corresponding original function possesses.

Since division by s corresponds to integration, division by $s^{1/2}$ corresponds to half-order integration. We ignore powers of $2\pi i$.

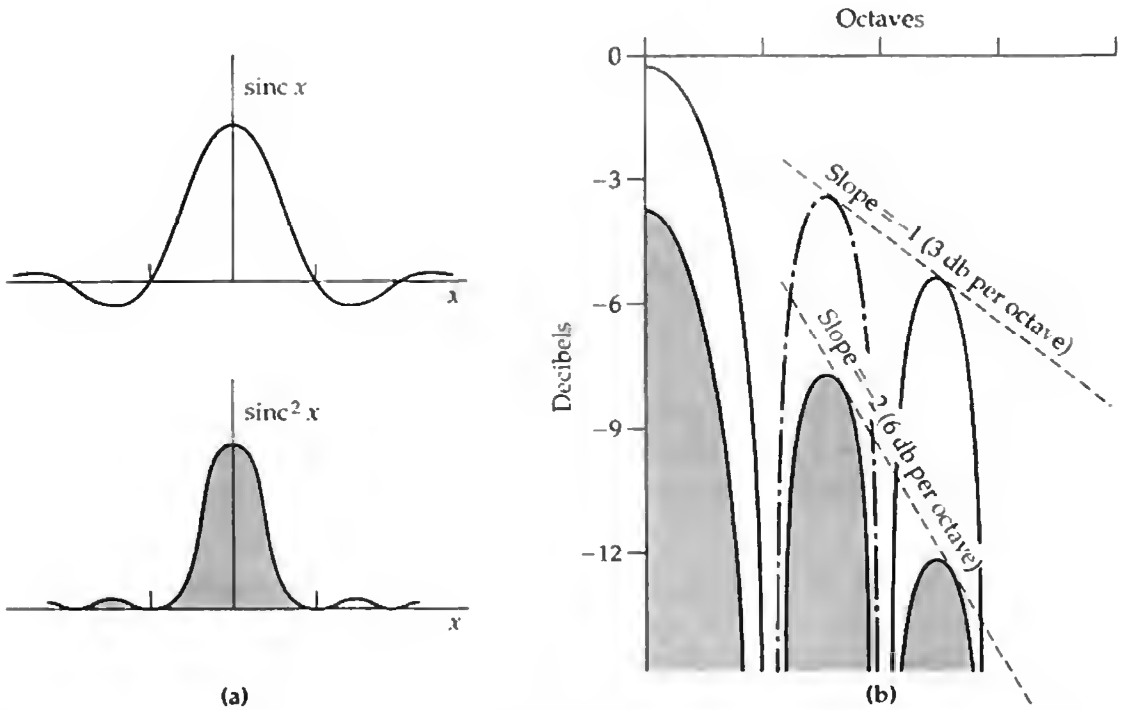


Fig. 8.7 **Asymptotic behavior as expressed on a double logarithmic plot in decibels per octave: (a) linear plot for comparison; (b) logarithmic plot (negative lobes shown broken).**

The examples tabulated in Table 8.2 show that for the purposes of asymptotic behavior there are indeed discontinuities intermediate in severity between jumps and corners. Thus a corner where a function changes its slope from horizontal to vertical (A) ranks midway between simple corners and simple jumps. A discontinuity (B) midway between a simple jump and an impulsive infinity proves to be less severe.

There are further subtleties to this matter. Thus $\delta(x)$ and x^{-1} prove to have the same kind of infinity at $x = 0$, and the infinite discontinuity of $\log x$, which is of the weakest possible kind, proves to be equivalent to a simple jump.

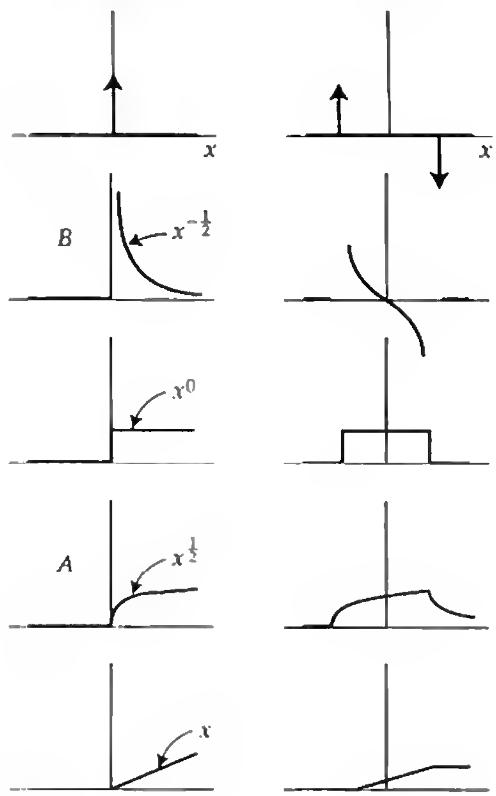
EQUIVALENT WIDTH

A convenient measure of the width of a function, for certain shapes of function, is

$$\frac{\int_{-\infty}^{\infty} f(x)\,dx}{f(0)},$$

TABLE 8.2
Fractional-order discontinuities

Asymptotic behavior of spectrum		$H(x)$	$\Pi(x)$
First derivative	s^0		
Half-order derivative	$s^{-\frac{1}{2}}$		
Original	s^{-1}		
Half-order integral	$s^{-\frac{3}{2}}$		
Integral	s^{-2}		



that is, the area of the function divided by its central ordinate. Expressed differently, the equivalent width of a function is the width of the rectangle whose height is equal to the central ordinate and whose area is the same as that of the function. This is illustrated in Fig. 8.8.

Examples can be drawn from many sources. In spectroscopy the equivalent width of a spectral line is defined as the width of a rectangular profile which has the same central intensity and the same area as the line. In antenna theory the effective beamwidth of an antenna has the character of an equivalent width (Chapter 15). Other examples occur in radiometry (Chapter 17). Some examples of functions, with accompanying rectangles illustrating their equivalent width, are shown in Fig. 8.9; all the examples have the same equivalent width.

If $f(0) = 0$, the equivalent width does not exist. For all cases, however, where $f(0) \neq 0$ and $\int_{-\infty}^{\infty} f(x)\,dx$ exists, there exists an "equivalent width," even though the character of equivalence in the spectroscopic sense is lost.

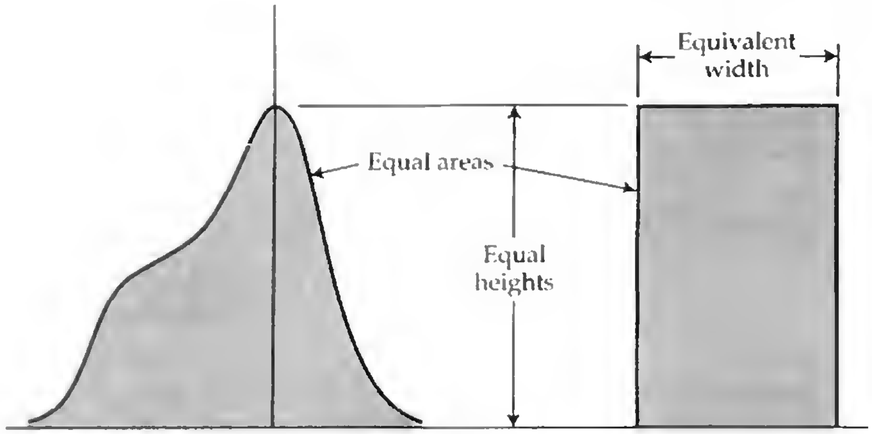


Fig. 8.8 The "equivalent width" of a function.

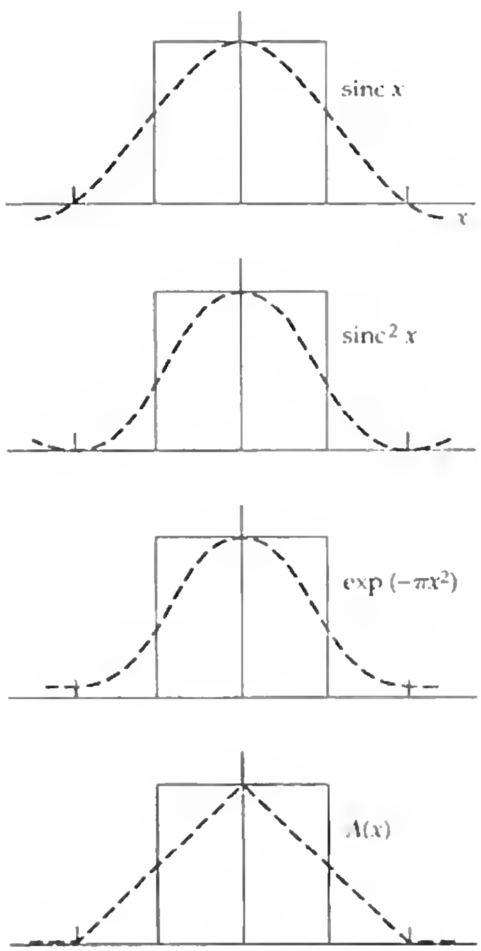


Fig. 8.9 Equivalent widths of some common distributions.

One expects that the wider a function is, the narrower its spectrum will be, but as functions of various shapes are contemplated, can we express this inverse relationship quantitatively? It turns out that a very precise relation exists. Whenever the equivalent width exists, it possesses the following elegant property.

The equivalent width of a function is equal to the reciprocal of the equivalent width of its transform; that is,

$$\frac{\int_{-\infty}^{\infty} f(x)\,dx}{f(0)} = \frac{F(0)}{\int_{-\infty}^{\infty} F(s)\,ds}.$$

This theorem is reminiscent of the similarity theorem, according to which the compression and expansion of a function and its transform behave reciprocally. The similarity theorem, however, is restricted to functions of a given shape. The present theorem says what quantity it is that behaves reciprocally, and it permits unlike functions to be ranked in order of that quantity.

Equivalent width is not always the best measure of width. For example, in expressing the bandwidth of a filter or the beamwidth of an antenna, it is common to adopt the "width to half power." According to this measure the second example illustrated above is narrower, for some purposes, than the first, which agrees with experience.

An interesting paradox arises when a localized turbulent motion spreads outward in space while the turbulence spectrum spreads out to higher wave numbers. In this case equivalent width is the wrong measure of spread.

The effective solid angle of an antenna radiation pattern (which is equal to 4π/directivity) is the "equivalent width" of the pattern, generalized to two dimensions.

Exercise. Investigate the behavior of equivalent width under convolution, beginning with some exploratory trials on Gaussian and rectangle functions of various relative widths.

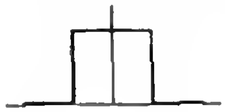
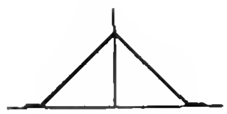
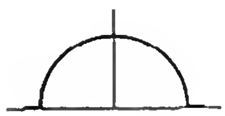
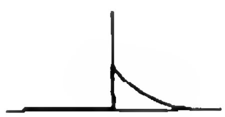
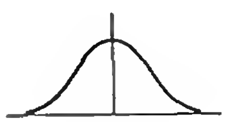
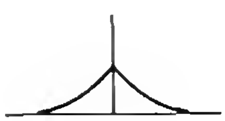
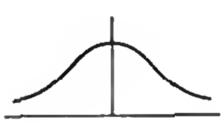
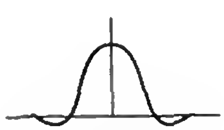
Table 8.3 lists equivalent widths for a number of common functions, together with autocorrelation widths, which are introduced in the next section. In this table the following abbreviations are used:

$$f \star f = \int_{-\infty}^{\infty} f(x+u) f(u)\,du$$

$$W_f = \frac{1}{f(0)} \int_{-\infty}^{\infty} f\,dx$$

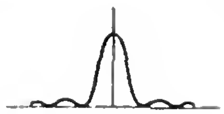
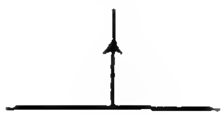
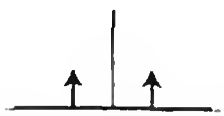
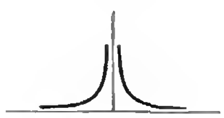
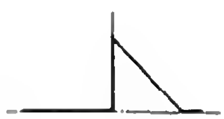
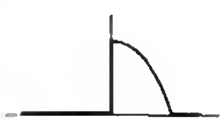
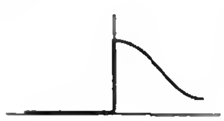
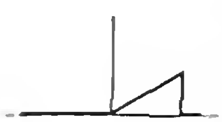
$$W_{f \star f} = \frac{1}{f \star f|_0} \int_{-\infty}^{\infty} (f \star f)\,dx$$

TABLE 8.3
Equivalent widths and autocorrelation widths

$f(x)$		$\int f\,dx$	$f(0)$	W_f	$f \star f$	$\int (f \star f)\,dx$	$f \star f\vert_0$	$W_{f\star f}$
$\Pi(x)$		1	1	1	$\Lambda(x)$	1	1	1
$\Lambda(x)$		1	1	1	$\Lambda * \Lambda$	1	$\frac{2}{3}$	$\frac{3}{2}$
$\cos \pi x \, \Pi(x)$		$2/\pi$	1	$2/\pi$	$\left\{\frac{1}{2}(1 - \lvert x\rvert) \cos \pi x + \frac{1}{2\pi} \sin \pi \lvert x\rvert\right\} \Pi\left(\frac{\pi}{2}\right)$	$4/\pi^2$	$\frac{1}{2}$	$8/\pi^2$
$e^{-x}H(x)$		1	1†	1	$\frac{1}{2}e^{-\lvert x\rvert}$	1	$\frac{1}{2}$	2
$e^{-\pi x^2}$		1	1	1	$2^{-\frac{1}{2}}e^{-\pi x^2/2}$	1	$2^{-\frac{1}{2}}$	$2^{\frac{1}{2}}$
$e^{-\lvert x\rvert}$		2	1	2	$(\lvert x\rvert + 1)e^{-\lvert x\rvert}$	4	1	4
$\dfrac{1}{1 + x^2}$		π	1	π	$\dfrac{2\pi}{4 + x^2}$	π^2	$\pi/2$	2π
sinc x		1	1	1	sinc x	1	1	1

†The value given is $f(0+)$.

TABLE 8.3
Equivalent widths and autocorrelation widths (cont'd)

$f(x)$	$\int f\,dx$	$f(0)$	W_f	$f \star f$	$\int (f \star f)\,dx$	$f \star f\|_0$	$W_{f \star f}$
$\mathrm{sinc}^2\, x$ 	1	1	1	$\frac{1}{\pi x^2}(1 - \mathrm{sinc}\, 2x)$	1	$\frac{2}{3}$	$\frac{3}{2}$
$\delta(x)$ 	1	∞	0	$\delta(x)$	1	∞	0
$\mathrm{II}(x)$ 	1	0	∞	$\frac{1}{4}\delta(x+1) + \frac{1}{2}\delta(x) + \frac{1}{4}\delta(x-1)$	1	∞	0
$2\,K_0(2\pi x)$ 				Fourier transform of $(1+x^2)^{-1}$	1	π	π^{-1}
$\Lambda(x)H(x)$ 	$\frac{1}{2}$	1†	$\frac{1}{2}$	$\frac{1}{6}\Pi\left(\frac{x}{2}\right)(\|x\|^3 - 3\|x\| + 2)$	$\frac{1}{4}$	$\frac{1}{3}$	$\frac{3}{4}$
$\cos \pi x\, \Pi(x)H(x)$ 	π^{-1}	1†	π^{-1}	$\frac{1}{4}(1 - 2\|x\| + \sin 2\pi\|x\|)$ $\cos^2 \pi x\, \Pi(x)$	π^{-2}	$\frac{1}{4}$	$4/\pi^2$
$\frac{H(x)}{1+x^2}$ 	$\pi/2$	1†	$\pi/2$	…	$\pi^2/4$	$\pi/4$	π
$x\Pi(x - \frac{1}{2})$ 	$\frac{1}{2}$	0	∞	$\frac{1}{6}\Pi\left(\frac{x}{2}\right)(\|x\|^3 - 3\|x\| + 2)$	$\frac{1}{4}$	$\frac{1}{3}$	$\frac{3}{4}$
$\frac{1}{1 + i2\pi ax}$	$\frac{1}{2}a$	1	$2a$	…	…	…	…
$\frac{1}{(1+x^2)^2}$	$\pi/2$	1	$\pi/2$	…	$\pi^2/4$	$5\pi/16$	$4\pi/5$

†The value given is $f(0+)$.

AUTOCORRELATION WIDTH

Another interesting "width" is the equivalent width of the autocorrelation function, which we shall call the autocorrelation width:

$$W_{f^* \star f} = \frac{\int_{-\infty}^{\infty} (f^* \star f)\, dx}{f^* \star f|_0} = \frac{\int f\, dx \int f^*\, dx}{\int f f^*\, dx}.$$

Autocorrelation width so defined does not necessarily refer to concentration about the origin. For example, the two functions shown in Fig. 8.10 would have the same autocorrelation width because they have the same autocorrelation function, but they would not have the same equivalent width. If a rectangular pulse were considered, then the idea of equivalent width could completely break down simply as a result of displacement along the axis of abscissas, because the central ordinate could fall to zero. This possibility is eliminated when one uses the equivalent width of the autocorrelation function, which is a maximum at the origin, never zero.

It is clear from the reciprocal property of the equivalent widths of a function and its transform that the autocorrelation width of a function is the reciprocal of *the equivalent width of its power spectrum*. Thus the idea of autocorrelation width breaks down if the equivalent width of the power spectrum loses meaning. This happens when the power spectrum has a zero central ordinate, which in turn means that the function (and its autocorrelation function) have zero area.

Hence, in relation to functions such as those used to represent wave packets which have little or no direct-current component, the autocorrelation width is not an appropriate measure of "extent" or duration.

The reciprocal of the autocorrelation width of the transform of a function is the equivalent width of the squared modulus of the function:

$$\frac{\int f f^*\, dx}{f(0) f^*(0)}.$$

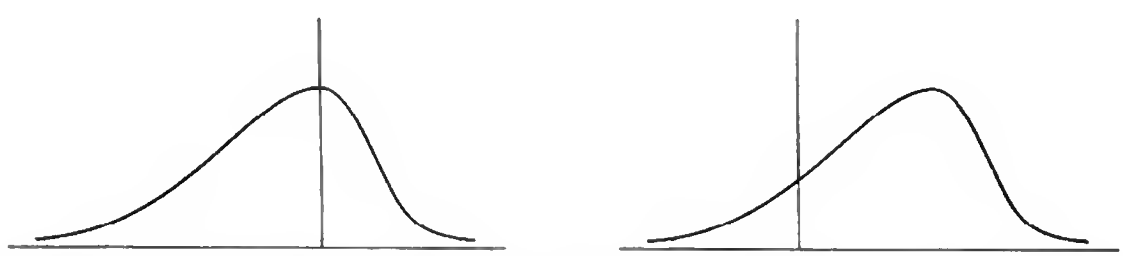


Fig. 8.10 **Functions with different equivalent widths but the same autocorrelation width.**

The advantage of autocorrelation width as a measure of dispersion comes from the removal of the sensitivity to the value of the central ordinate exhibited by the equivalent width. Two functions that have the same power spectrum, but whose spectral components are slid about into any relative phase, have the same autocorrelation function. Therefore the quantity $W_{f\star f}$ copes with (1) displacement of the function away from the origin and (2) internal phase interference of the spectral components, which happens to produce a small central ordinate. However, of two functions with the same power spectrum, one can be narrow and another wide. For example, the Fresnel diffraction radiation fields on planes parallel to an illuminated aperture have the same autocorrelation, but on any intuitive view the width of the illumination increases with distance from the source aperture.

Difficulty (1) can be removed by considering equivalent widths about the centroid. Difficulty (2) indicates that (*a*) internal phase relations of the components cannot be ignored and that (*b*) a width measure hinging on some one ordinate will always be sensitive as a consequence.

This critical approach to width measures should not obscure the fact that equivalent width and autocorrelation width both occur as the appropriate quantity in particular physical circumstances. Thus it would be pointless to criticize the sensitive dependence of antenna directivity on one ordinate of the radiation pattern.

A situation where the autocorrelation width occurs is rectification of white noise passed through a filter. It is well known qualitatively that the average interval between effectively independent output values is inversely proportional to the "width" of the pass band. In fact it is the autocorrelation width of the power pass characteristic for which this statement is quantitatively accurate.

An important property of the autocorrelation width which fits it for many applications is its invariance under shuffling (see Problems).

MEAN-SQUARE WIDTHS

Since the equivalent width and concentration just described do not fill all needs, other measures of width are also encountered. For example, the mean-square value of x defined by

$$\langle x^2 \rangle = \frac{\int_{-\infty}^{\infty} x^2 f(x)\, dx}{\int_{-\infty}^{\infty} f(x)\, dx}$$

is widely appropriate as a width measure, as is also the variance in functions not centered on the origin of x:

$$\begin{aligned}\langle (x - \langle x \rangle)^2 \rangle &- \langle x^2 \rangle - \langle x \rangle^2 \\ &= \frac{F''(0)}{-4\pi^2 F(0)} + \frac{1}{4\pi^2}\left[\frac{F'(0)}{F(0)}\right]^2.\end{aligned}$$

These measures break down when the area of the function is zero and are therefore inappropriate for measuring the width of an oscillatory signal or wave packet. To cope with this case we think in terms of the energy density and the centroid and variance of the energy distribution. We shall refer to the mean-square departure from the centroid of $|f(x)|^2$ as the variance of the squared modulus, and it is given by

$$(\Delta x)^2 = \frac{\int_{-\infty}^{\infty} x^2|f(x)|^2\,dx}{\int_{-\infty}^{\infty}|f(x)|^2\,dx} - \left[\frac{\int_{-\infty}^{\infty} x|f(x)|^2\,dx}{\int_{-\infty}^{\infty}|f(x)|^2\,dx}\right]^2.$$

This may seem an elaborate expression for the simple physical idea of duration or bandwidth of a signal packet, but it behaves more reasonably for many purposes than the simpler measures considered before.

Of course, the square of a given function may not have a finite variance; for example, sinc x and $(x^2 + 1)^{-\frac{1}{2}}$, two concentrated peaked functions having reasonable "equivalent widths" and "autocorrelation widths," show up as infinitely broad on a variance basis. Caution is therefore necessary before assuming that widths based on variance will have any relation to intuitive ideas of width in particular cases.

In the rigorous expression of the uncertainty relation referred to below, the squared uncertainties are variances of squared moduli. Suppose an observed quantity had the form sinc x, as would happen if an impulsive wave were sampled with receiving equipment of limited high-frequency response, or if an infinite plane wave were sampled through an aperture of limited spatial extent. Then that spread which is subject to the quantitative uncertainty relation would evidently be by no means as well in agreement with one's intuitive notion of the spread as it is in the case of a Gaussian packet $\exp(-ax^2)\sin x$.

Table 8.4 lists the centroid and the mean-square abscissa for a variety of functions.

SAMPLING AND REPLICATION COMMUTE

When a function $f(x)$ defined for values of x from $-\infty$ to ∞ is subjected to the operation of regular sampling at unit interval the result is the sample set $\{\ldots \quad f(x+2) \quad f(x+1) \quad f(x) \quad f(x-1) \quad f(x-2) \quad \ldots\}$. The result of replication of $f(x)$ at interval X is the function $\ldots + f(x+2X) + f(x+X) + f(x) + f(x-X) + f(x-2X) \ldots$. Sometimes, as with $f(x) = (1+x^2)^{-1}$, the replication sum does not exist; the following is not applicable to such functions.

If we replicate a function, and then sample, the result is a doubly infinite entity

TABLE 8.4
Centroid and mean-square abscissa

$f(x)$	$\langle x\rangle$	$\langle x^2\rangle$	$F''(0)$	$f''(0)$	$\Delta(x)^2$
$\Pi(x)$	0	$\frac{1}{12}$	$-\pi^2/3$	0	$\frac{1}{12}$
$\Lambda(x)$	0	$\frac{1}{6}$	$-2\pi^2/3$	∞	$\frac{1}{4}$
$\operatorname{sinc} x$	0	osc	0	$-\pi^2/3$	∞
$\operatorname{sinc}^2 x$	0	∞	∞	$-2\pi^2/3$	...
$H(x)$	∞	∞	∞	∞	∞
$e^{-x}H(x)$	1	2	$-8\pi^2$	∞	$\frac{1}{4}$
$e^{-\lvert x\rvert}$	0	4	$-16\pi^2$	∞	1
$\dfrac{1}{1+x^2}$	0	∞	∞	-2	...
$\cos \pi x \Pi(x)$	0	$\dfrac{\pi^2-8}{4\pi^2}$	$-\dfrac{2}{\pi}(\pi^2-8)$	$-\pi^2$	...
$(1-x^2)\Pi\left(\dfrac{x}{2}\right)$	0	$\frac{4}{15}$	$-\dfrac{16\pi^2}{15}$	-2	...
$e^{-\pi x^2 a^2}$	0	$\dfrac{1}{2\pi a^2}$	$-\dfrac{2\pi}{a^3}$	$-2\pi a^2$	$\dfrac{1}{4\pi a^2}$
$e^{-\pi x^2 a^2}\cos \omega x$	0	$\dfrac{1}{2\pi}-\left(\dfrac{\omega}{2\pi a}\right)^2$	$-\dfrac{2\pi}{a}\left(1-\dfrac{\omega^2}{2\pi a^2}\right)e^{-\pi(\frac{\omega}{2\pi a})^2}$	$-(\omega^2+2\pi a^2)$	...
$\text{Ш}(x)$	0	$\frac{1}{4}$	$-\pi^2$	0	∞
$\operatorname{sech} \pi x$	0	$\frac{1}{4}$	$-\pi^2$	$-\pi^2$	...
$xe^{-x}H(x)$	2	6	$-24\pi^2$	∞	...

$$\begin{bmatrix} \cdots + f(x+2) + \cdots \\ \cdots + f(x+1+2X) + f(x+1+X) + f(x+1) \\ \qquad + f(x+1-X) + f(x+1-2X) + \cdots \\ \cdots + f(x+2X) + f(x+X) + f(x) + f(x-X) + f(x-2X) + \cdots \\ \cdots + f(x-1+2X) + f(x-1+X) + f(x-1) \\ \qquad + f(x-1-X) + f(x-1-2X) + \cdots \\ \cdots + f(x-2) + \cdots \end{bmatrix}$$

shown here in the form of a column matrix, or vector, whose elements are the resulting sample set.

If on the other hand we propose to sample first, and then replicate, it is necessary first to extend the meaning of replication (introduced above as an operation on a function) so as to be applicable to a sample set. First restrict the repli-

cation interval X to an integer. Let the sample set be f_i for $-\infty < i < \infty$. Then the replicated sample set r_j will be defined by

$$r_j = \cdots f_{j+2X} + f_{j+X} + f_j + f_{j-X} + f_{j-2X} + \cdots,$$

(provided the sums exist). Under these definitions, the replicated sample set is the same as the sampled replication.

An alternative statement of this result uses the equivalence between a sample set a_i and the generalized function $\sum_i a_i \delta(x - i)$. Sampling at unit interval then consists of multiplication by $\mathrm{III}(x)$, while replication at interval X consists of convolution with $X^{-1}\mathrm{III}(x/X)$. In terms of the sampling operation $\mathrm{III}(x) \times$ and the replication operator $X^{-1}\mathrm{III}(x/X) *$, the commutativity property for sampling and replication then reads

$$X^{-1}\mathrm{III}(x/X) * [\mathrm{III}(x) \times f(x)] = \mathrm{III}(x) \times [X^{-1}\mathrm{III}(x/X) * f(x)].$$

Referring to Fig. 10.17 may help to illustrate this relation.

SOME INEQUALITIES

The discontinuities, cusps, or other sharp behavior characteristics of a function reveal the presence of high frequencies in its spectrum; in fact a quantitative measure of the sharpness of the behavior is linked with the decay of the spectrum in another section. Now if a function is quite continuous, with all derivatives continuous, there is still a limit to the slope, or curvature, which it can have.

If such a function undergoes a large change in a short interval of abscissa as in Fig. 8.11 (that is, it has a very large slope in this interval), it may for some practical purposes be effectively discontinuous. Its spectrum will indeed be found to behave appropriately. Asymptotic behavior appropriate to the continuous function will ultimately set in, but only at high frequencies beyond the range of interest.

Upper limits to ordinate and slope. Just how steep can the slope of a function be? Consider first the problem of how great a function can become. We can say that

$$|f(x)| \le \int_{-\infty}^{\infty} |F(s)|\, ds$$

as a direct consequence of the relation

$$f(x) = \int_{-\infty}^{\infty} F(s) e^{i2\pi xs}\, ds.$$

On the complex plane of $f(x)$ a Fourier integral for a particular value of x is illustrated in Fig. 8.12. For other values of x the locus would be different but the

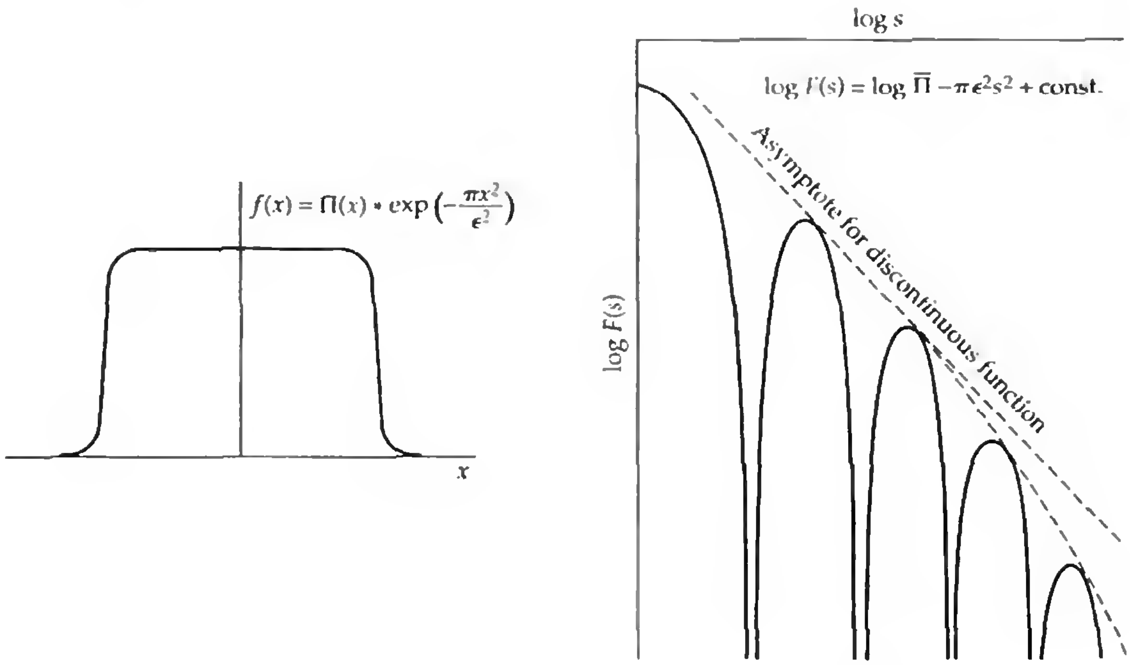


Fig. 8.11 A continuous function with all continuous derivatives, which for certain purposes is effectively discontinuous, and its transform.

length of the arc, which is equal to

$$\int_{-\infty}^{\infty} |F(s)|\, ds,$$

must remain the same. This arc length is the maximum value which $|f(x)|$ could assume; and it would do so if there were a value of x for which the arc extended into a straight line. In other words, the maximum value of a function cannot exceed the resultant of all its components at that point, if it exists, where they all come into phase together.

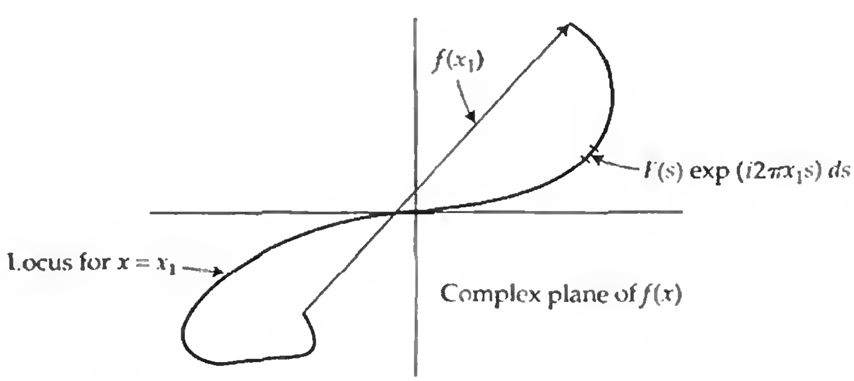


Fig. 8.12 Locus on the complex plane of $f(x)$.

The slope of the component $F(s)\exp(i2\pi xs)ds$ is $2\pi isF(s)\exp(i2\pi xs)ds$, and by considering the possibility of all components having their maximum slopes together, we have

$$|f'(x)| \leq 2\pi \int_{-\infty}^{\infty} |sF(s)|\, ds.$$

The two inequalities we have derived both stem from the geometrical phenomenon that an arc joining two points is equal to or greater than the chord. This in turn can be further distilled and depends ultimately on the property that two sides of a triangle are together greater than or equal to the third. In vector notation

$$|\mathbf{A} + \mathbf{B}| \leq |\mathbf{A}| + |\mathbf{B}|.$$

It appears that inequalities are often reducible to one or the other of a few geometrical inequalities. This particular one is referred to as Bessel's inequality.

Schwarz's inequality. If two real functions $f(x)$ and $g(x)$ are defined in the interval $a < x < b$, then

$$\left[\int_a^b f(x)g(x)\, dx\right]^2 \leq \int_a^b [f(x)]^2\, dx \int_a^b [g(x)]^2\, dx.$$

A related vector inequality states that the scalar product of two vectors cannot exceed the product of their absolute values; thus

$$\mathbf{A} \cdot \mathbf{B} \leq AB.$$

If F and G are complex functions of x, we have the Schwarz inequality in the form most suited to the present work:

$$\left[\int (F^*G + FG^*)\, dx\right]^2 \leq 4 \int FF^*\, dx \int GG^*\, dx.$$

To prove Schwarz's inequality, let ϵ be a real constant. Then

$$0 < \int (F + \epsilon G)(F + \epsilon G)^*\, dx,$$

because the integrand is a positive quantity (unless by accident F happens to be proportional to G, in which case the value of ϵ that makes $F + \epsilon G$ identically equal to zero must be avoided). By expanding the above inequality we can write

$$0 < \int FF^*\, dx + \epsilon \int (F^*G + FG^*)\, dx + \epsilon^2 \int GG^*\, dx.$$

The right-hand side is a quadratic expression in ϵ^2; call it $c + b\epsilon + a\epsilon^2$. The condition that it should not reduce to zero for any real value of ϵ is $b^2 - 4ac < 0$.

When the Schwarz inequality is applied, definite limits may be assigned to the integrals. For example, in connection with the uncertainty relation infinite limits are adopted. The inequality holds, however, for any limits.

THE UNCERTAINTY RELATION

It is well known that the bandwidth-duration product of a signal cannot be less than a certain minimum value. This is essentially a mathematical phenomenon bound in with the interdependence of time and frequency, which prevents arbitrary specification of signals on the time-frequency plane. One may specify arbitrary functions of time or arbitrary spectra, but not both together. Thus a finite area of the time-frequency plane can contain only a finite number of independent data. The bandwidth-duration product of a signal cannot be less than the value for an elementary signal containing only one datum. Such a signal may be brief and wideband or quasi-monochromatic and persistent.

Since the equivalent widths of a function and its transform are reciprocals, it follows that

$$\text{Equivalent duration} \times \text{equivalent bandwidth} = 1.$$

The product of the autocorrelation widths is deducible from a previous section as

$$\frac{|f(0)|^2 \left| \int f\,dx \right|^2}{\left[\int |f|^2\,dx \right]^2},$$

and if we deal in mean-square widths,

$$\langle x^2 \rangle \langle s^2 \rangle = -\frac{f''(0) \int x^2 f\,dx}{4\pi^2 f(0) \int f\,dx} = \frac{f''(0)F''(0)}{16\pi^4 f(0)F(0)}.$$

Neither of these last two width products is a constant, and the second can range from zero to infinity.

Now consider the energy functions. Let $(\Delta x)^2$ be the variance of $|f(x)|^2$ and $(\Delta s)^2$ be the second moment of $|F(s)|^2$, to obtain the usual expression of the uncertainty relation, namely,

$$\Delta x\,\Delta s \geq \frac{1}{4\pi}.$$

Proof of uncertainty relation. In this section all integrals are to be taken between infinite limits. We need the theorem

$$\int f' f'^{*}\,dx = 4\pi^2 \int s^2 F F^{*}\,ds,$$

which may be verified by combining the derivative theorem with Rayleigh's theorem.

We also need the Schwarz inequality in the form

$$4\int ff^*\,dx \int gg^*\,dx \geq \left|\int (f^*g + fg^*)\,dx\right|^2,$$

and the formula for integration by parts between infinite limits in the form

$$\left|\int xf'\,dx\right| = \left|\int f\,dx\right|.$$

Now, taking $f(x)$, and thus also $F(s)$, to be centered on its centroid,

$$
\begin{aligned}
(\Delta x)^2(\Delta s)^2 &= \frac{\int x^2 ff^*\,dx \int s^2 FF^*\,ds}{\int ff^*\,dx \int FF^*\,ds} \\
&= \frac{\int xf.xf^*\,dx \int f'f'^*\,dx}{4\pi^2\left(\int ff^*\,dx\right)^2} \\
&\geq \frac{\left|\int (xf^*.f' + xf.f'^*)\,dx\right|^2}{16\pi^2\left(\int ff^*\,dx\right)^2} \\
&= \frac{\left|\int x\frac{d}{dx}(ff^*)\,dx\right|^2}{16\pi^2\left(\int ff^*\,dx\right)^2} \\
&= \frac{\left|\int ff^*\,dx\right|^2}{16\pi^2\left(\int ff^*\,dx\right)^2} \\
&= \frac{1}{16\pi^2}.
\end{aligned}
$$

Therefore $\Delta x\,\Delta s \geq \dfrac{1}{4\pi}$.

Example of uncertainty relation. Take $f(x)$ to be a voltage $V(t)$ whose Fourier transform is $S(f)$. We select $V(t)$ so that its spectrum is real for the purpose of illustration but in general $S(f)$ is complex. Voltages or the fields of electromagnetic waves, on the other hand, can only be real. The signal energy distributions and the energy spectrum are shown in Fig. 8.13, together with Δt and Δf the root-

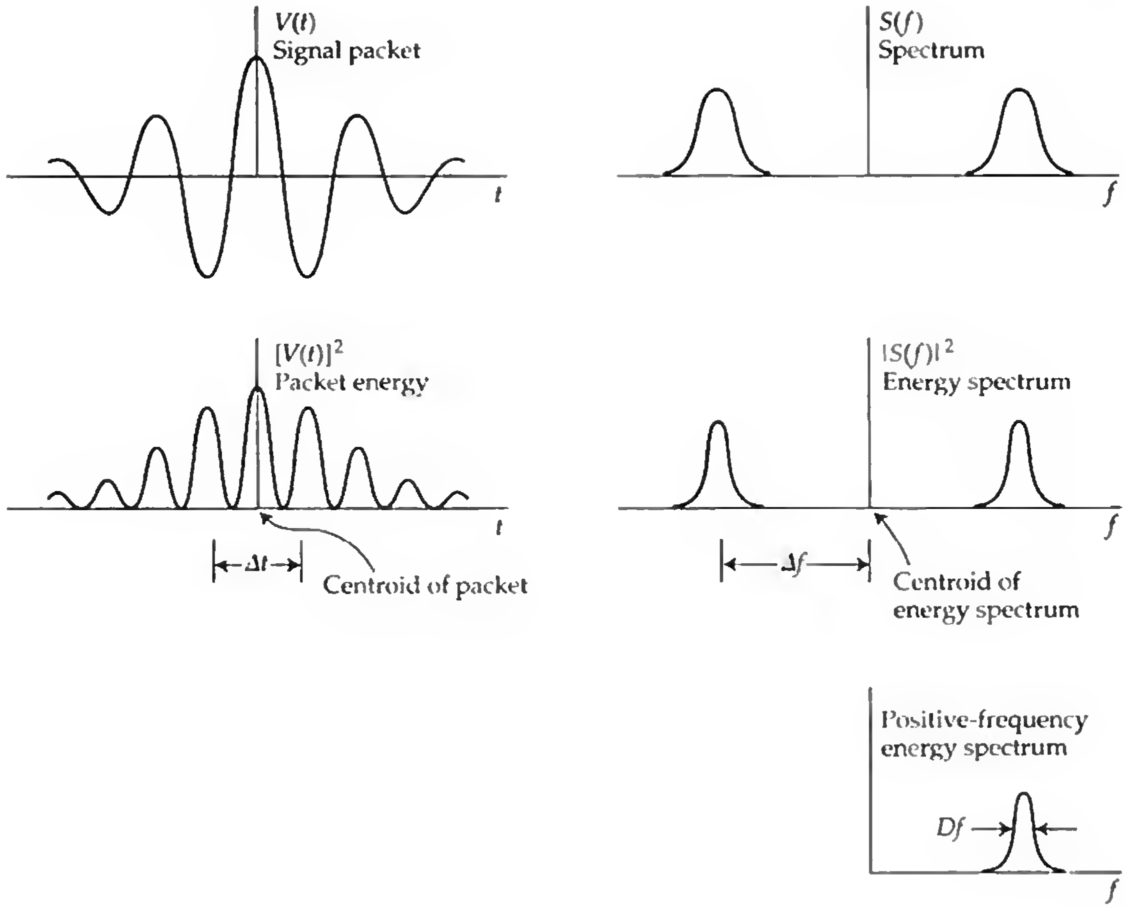


Fig. 8.13 The uncertainty relation.

mean-square departures from the centroids of the two energy distributions. The uncertainty relation says that the product $\Delta t\ \Delta f$ must exceed or equal $(4\pi)^{-1}$. In terms of angular frequency ω,

$$\Delta t\ \Delta\omega \geq \frac{1}{2}.$$

Let $f(x) = \exp(-\pi a^2 x^2)$. Then $ff^* = \exp(-2\pi a^2 x^2)$, for which the variance of x is $1/4\pi a^2$, and so $\Delta x = (4\pi a^2)^{-\frac{1}{2}}$. Similarly, for the transform, $FF^* = [a^{-1}\exp(-\pi s^2/a^2)]^2$ with variance $a^2/4\pi$ and $\Delta s = (4\pi/a^2)^{-\frac{1}{2}}$. Hence, for this case,

$$\Delta x\ \Delta s = \frac{1}{4\pi},$$

which is the minimum value permitted by the uncertainty relation in its quantitative form.

However, we can make a much stronger statement. Suppose that a radar transmitter emits a 1-microsecond pulse at a frequency of 10,000 megahertz. Then the uncertainty principle says that

$$10^{-6} \geq \frac{1}{4\pi \times 10^{10}}.$$

Now let the positive-frequency power spectrum defined for $f > 0$ by $2|S(f)|^2$ have a width Df defined by

$$(Df)^2 = \frac{\int_0^\infty |f - \langle f \rangle|^2 |S(f)|^2 \, df}{\frac{1}{2}\int_0^\infty [s(t)]^2 \, dt}.$$

We know from experience with radio receivers that

$$\Delta t \sim \frac{1}{Df},$$

since in the case of a 1-microsecond pulse a receiver bandwidth of 1 megahertz is an optimum choice for detection. The uncertainty relation thus sets a lower limit to the duration-bandwidth product, which in the case illustrated fails by a factor of 10^5 to be practical.

While the positive-frequency energy spectrum is not quite as amenable to mathematical handling as $S(f)$, the conventional spectrum in terms of which Δf is defined, it is closely connected with what would be observed if the energy spectrum were explored experimentally with tunable resonators equipped for energy detection.

The next step is to consider the product

$$\Delta t \, Df,$$

which can be shown to agree with experience in cases such as that quoted but which is not subject to the lower limit of $(4\pi)^{-1}$ applicable to $\Delta t \, \Delta f$. Whether the product $\Delta t \, Df$ has a lower limit has been studied by Uffink and Hilgevoord (1985, 1988). As an example that does not observe the limit consider the case of $s(t) = \exp(-\pi t^2) \cos 2\pi f_0 t$ and its Fourier transform $S(f) = \frac{1}{2}\exp[-\pi(f + f_0)^2 + \frac{1}{2}\exp[-\pi(f - f_0)^2]$, where $f_0 = 0.218$ as illustrated in Fig. 8.14. The heavy bars show Δt and Df. Their product equals $0.400 \times (1/4\pi)$, which is well below the quantum limit $\Delta t \, \Delta f$.

THE FINITE DIFFERENCE

We define the finite difference of $f(x)$, taken over the interval a, to be

$$\Delta_a f(x) = f(x + \tfrac{1}{2}a) - f(x - \tfrac{1}{2}a).$$

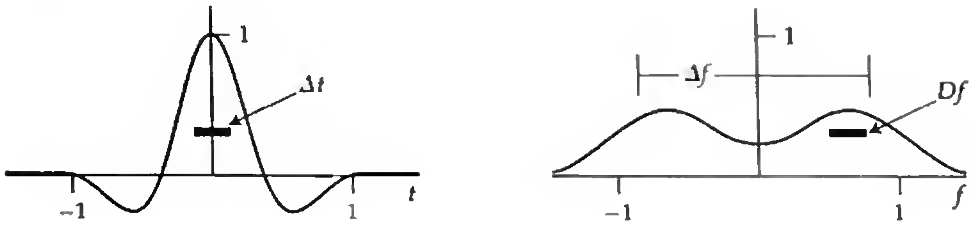


Fig. 8.14 Illustrating the discrepancy between Δf, which is subject to a lower limit set by $\Delta f \Delta t \leq 1/4\pi$, and Df, the width of the positive-frequency spectrum, which can be smaller.

The finite difference is often encountered in connection with functions that are tabulated at discrete intervals and so is defined only at discrete values of x. However, we may also take the finite difference of a function that is defined for all x, and in that case the finite difference is also a function of the continuous variable x.

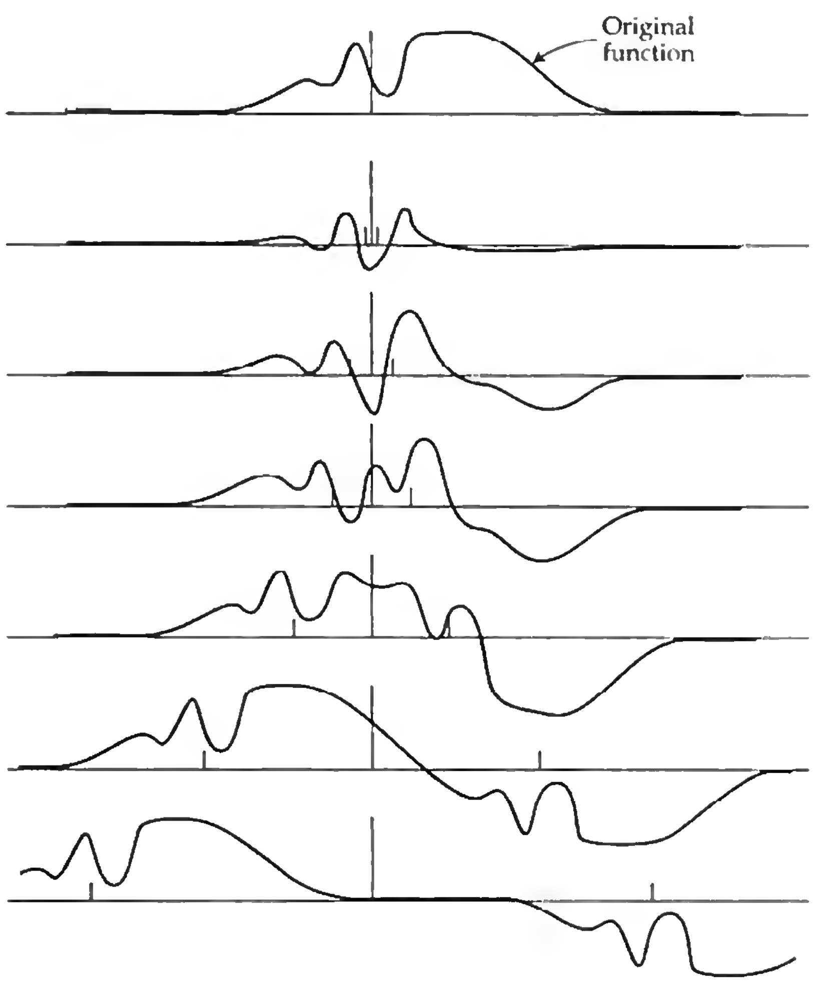


Fig. 8.15 Finite differences taken over different intervals.

When the differencing interval a is small, then $\Delta_a f(x)$ is small but approaches proportionality to the derivative; when a is large, $\Delta_a f(x)$ may separate out into a displaced $f(x)$ followed by an inverted $f(x)$. These possibilities are illustrated in Figs. 8.15 and 8.16.

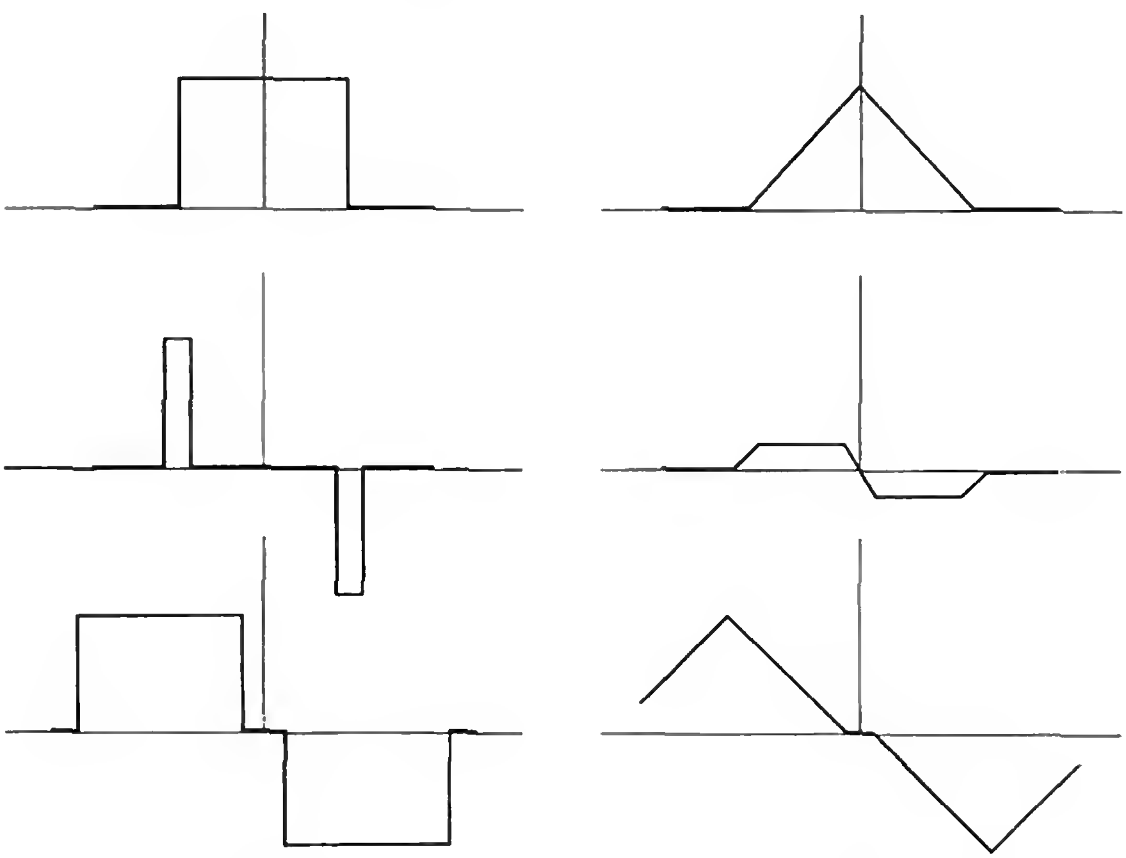
Since finite differencing in the x domain corresponds to a simple multiplying operation in the transform domain, it pays to be alert to the possibility of expressing functions as finite differences.

Exercise. Deduce the original functions of which the examples in Fig. 8.17 are the finite differences (if possible). Can you say whether your solutions are unique?

Exercise. Discuss the question of whether $\Delta_a f(x)$ is an odd function.

The process of taking the finite difference will be recognized as the equivalent of convolution; that is,

$$\Delta_1 f(x) = 2\,\mathrm{I}_{\mathrm{I}}(x) * f(x).$$



(a) (b)

Fig. 8.16 **Finite differences taken over short and long intervals of (a) rectangle function, (b) triangle function.**

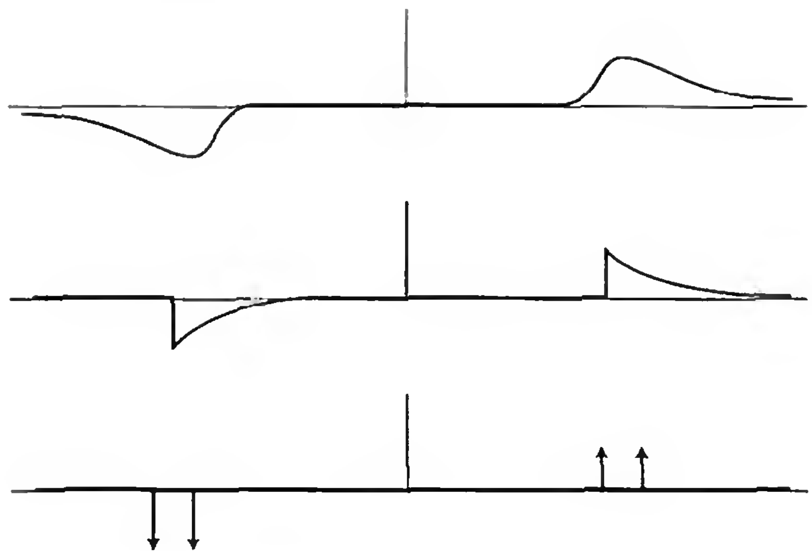


Fig. 8.17 **Functions to be expressed as finite differences.**

The Fourier transform of the odd impulse pair has already been discussed; hence we know that

$$\frac{1}{|a|}\,\mathrm{I}_{\mathrm{I}}\left(\frac{x}{a}\right) \supset i \sin \pi a s.$$

Consequently we have, by the convolution theorem, the following precise connection between the two domains.

If $\Delta_a f(x)$ is the finite difference of $f(x)$, taken over the interval a, then the Fourier transform of $\Delta_a f(x)$ is $2i \sin \pi a s\, F(s)$.

Finite differencing thus corresponds, in the transform domain, to multiplication by $2i \sin \pi a s$.

Examples given in Fig. 8.18 show how differencing over a wide interval incurs multiplication of the transform by a high-frequency sinusoid, but differencing over a small interval incurs multiplication by a low-frequency sinusoid. For a *very* small interval—for example, one shorter than the scale of the finest structure in $f(x)$—the sinusoid becomes effectively a linear function of s.

The connection with differentiation may be expressed, in the x domain by

$$\frac{d}{dx} \equiv \lim_{a \to 0} \frac{\Delta_a}{a},$$

and in the transform domain by

$$\lim \frac{2i \sin \pi a s}{a} = i2\pi s.$$

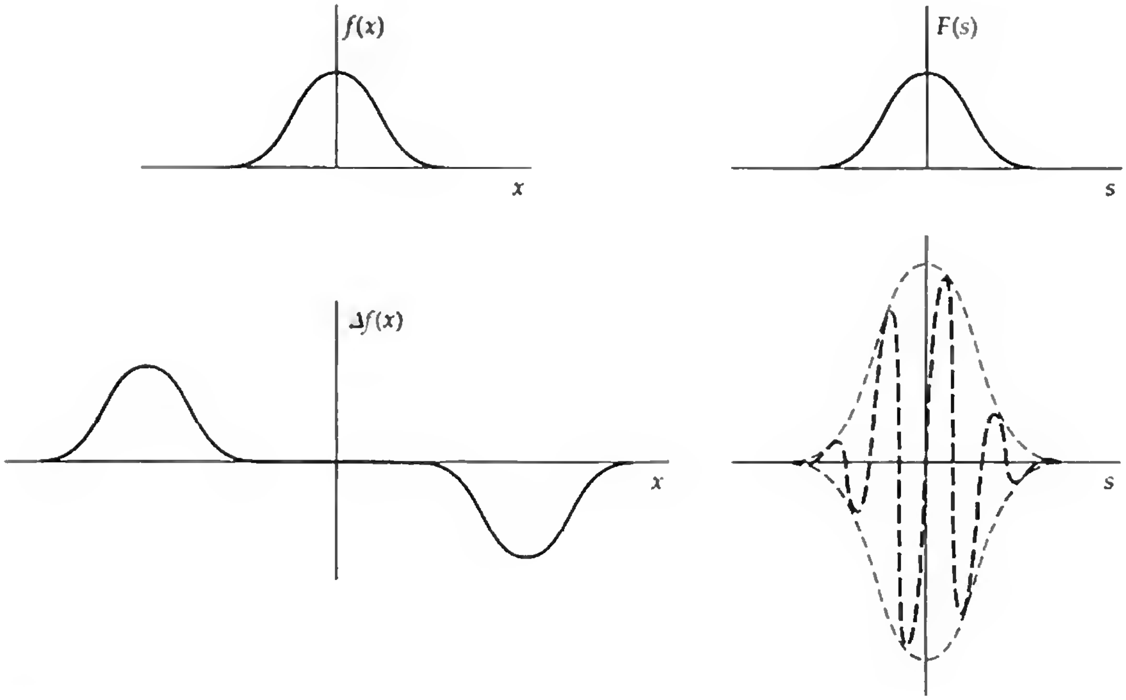


Fig. 8.18 **Finite differencing corresponds to multiplication of the transform by a sinusoid.**

The second difference $\Delta^2_{aa}f(x)$ is defined by

$$\Delta^2_{aa}f(x) = \Delta_a[\Delta_a f(x)] = f(x + a) - 2f(x) + f(x - a).$$

Since $\Delta^2_{aa}f(x)$ is obtained by taking the finite difference of the finite difference, it follows that it corresponds in the transform domain to two successive multiplications by $2i \sin \pi as$. Hence the Fourier transform of $\Delta^2 f(x)$ is $-4 \sin^2 \pi as\, F(s)$.

Geometrical interpretations of the finite differences are given in Fig. 8.19. The first difference is proportional to the average slope and the second difference to the average curvature over finite intervals.

Exercise. Show that $d^2/dx^2 \equiv a^{-2} \lim_{a\to 0} \Delta^2$.

RUNNING MEANS

In the reduction of meteorological data, such as rainfall data, the unimportant day-to-day fluctuations are customarily smoothed out to reveal the meaningful

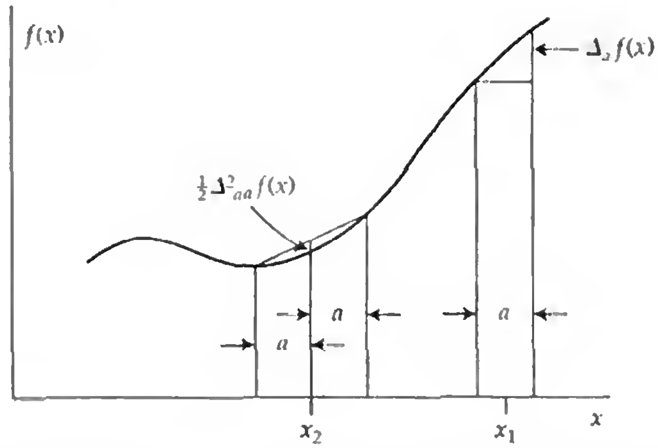


Fig. 8.19 **Geometrical interpretation of the first difference of $f(x)$ at $x = x_1$ and the second difference at $x = x_2$.**

seasonal trend by taking running means. The running mean of $f(x)$ over the interval a is defined by

$$\frac{1}{a}\int_{x-\frac{1}{2}a}^{x+\frac{1}{2}a} f(x')\,dx'.$$

It is quite general practice, when recording data of any sort automatically, to smooth out unwanted rapid fluctuations; for example, a signal to be fed to a recording milliammeter is often passed through a smoothing filter. The running mean is one kind of smoothed value; it is generated by electronic counters that integrate an input voltage and print out the result at the end of fixed time intervals. The usual method of integrating a voltage is to apply the voltage to a voltage-to-frequency converter and to count the cycles of the voltage controlled oscillation over a given time interval.

It is evident that the running mean may be expressed in the form of a convolution with a rectangle function of base a and unit area:

$$a^{-1}\Pi\left(\frac{x}{a}\right) * f(x).$$

Since the Fourier transform of $a^{-1}\Pi(x/a)$ is sinc as, the convolution theorem says that *the transform of the running mean of $f(x)$, taken over the interval a, is* sinc as *$F(s)$*. Thus the process of taking running means corresponds, in the transform domain, to multiplication by sinc as.

Exercise. Interpret the properties

$$\lim_{a\to 0} \operatorname{sinc} as = 1 \qquad \text{and} \qquad \lim_{a\to 0} \operatorname{sinc} as = a^{-1}\delta(s)$$

in terms of running means.

Running means of higher order result from successive applications of the same process; thus the second-order running mean

$$a^{-1}\Pi\left(\frac{x}{a}\right) * a^{-1}\Pi\left(\frac{x}{a}\right) * f(x)$$

has a Fourier transform

$$\operatorname{sinc}^2 as\, F(s).$$

Recognizing $\operatorname{sinc}^2 as$ as the transform of $a^{-1}\Lambda(x/a)$, we note that the second-order running mean may be regarded as a *weighted* running mean generated by convolution with $a^{-1}\Lambda(x/a)$.

A number of successive running means of an original rectangle function, shown in Fig. 8.20 together with the successive transforms, illustrate the relation

$$a^{-1}\Pi\left(\frac{x}{a}\right) * \ldots n \text{ times} \ldots * a^{-1}\Pi\left(\frac{x}{a}\right) \supset \operatorname{sinc}^n as.$$

CENTRAL-LIMIT THEOREM

If a large number of functions are convolved together, the resultant may be very smooth (Fig. 8.20), and as the number increases indefinitely, the resultant may approach Gaussian form. The rigorous statement of this tendency of protracted convolution is the central-limit theorem. Let us illustrate with an example.

In Fig. 8.20 two sequences of profiles are shown, each approaching Gaussian form in the central part as $n \to \infty$. Over a small range of s around $s = 0$ the top right-hand profile sinc s is approximated by the parabolic curve $1 - ms^2$, and therefore the nth profile is approximated by $(1 - ms^2)^n$. Now

$$\begin{aligned}(1 - ms^2)^n &= \left[1 - nms^2 + \frac{n(n-1)}{1.2} m^2s^4 - \ldots\right] \\ &= \left[1 - nms^2 + \frac{n^2m^2s^4}{1.2} - \ldots\right] + \text{remainder} \\ &= e^{-nms^2} + \text{remainder}.\end{aligned}$$

Keeping s fixed and letting n tend to infinity we see that we have a Gaussian term which continually gets narrower, its width varying as $n^{-\frac{1}{2}}$, plus a remainder of diminishing relative importance. The central curvature of the right-hand profile increases proportionally to n.

The Gaussian term has a Gaussian transform

$$\left(\frac{\pi}{nm}\right)^{\frac{1}{2}} e^{-\pi^2x^2/nm},$$

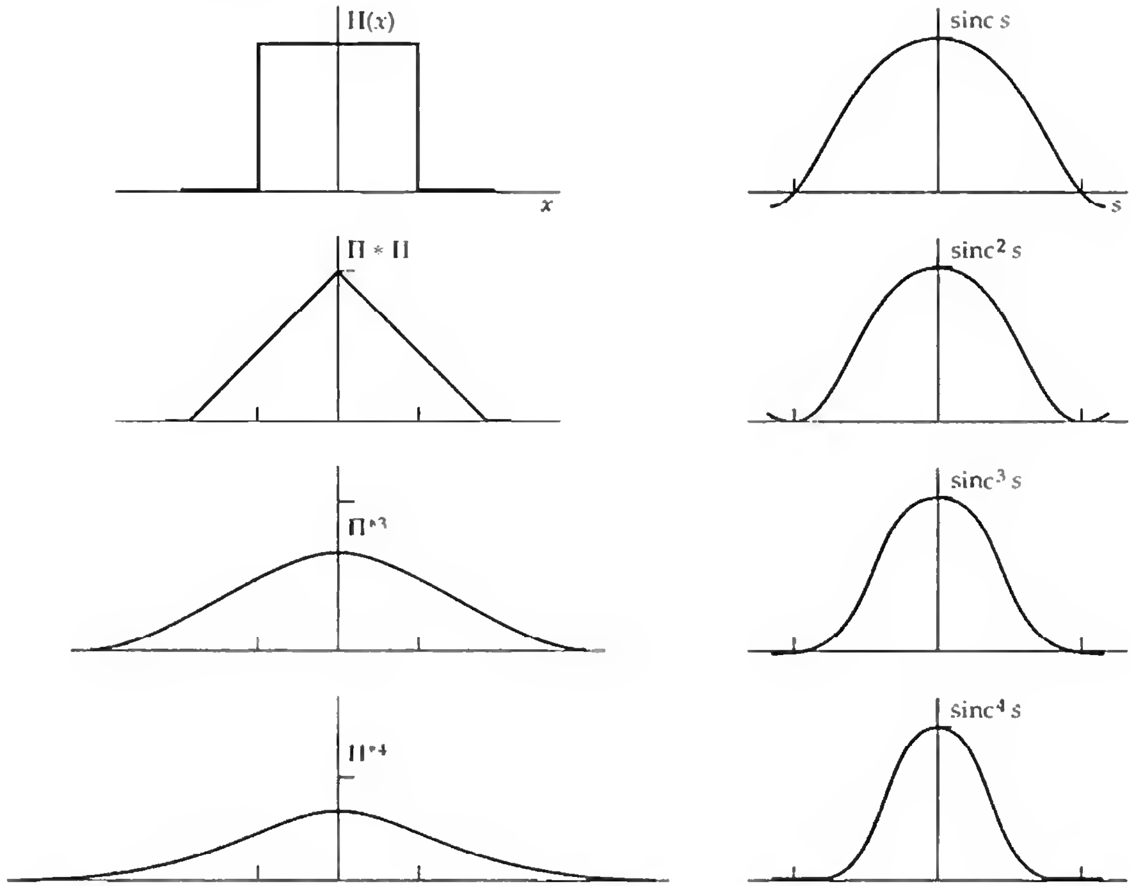


Fig. 8.20 **Successive running means and their transforms.**

which continually gets wider, the variance increasing proportionally to n; this, of course, is a consequence of the property that variances add under convolution.

The interesting phenomenon is the tendency toward Gaussian form under successive self-convolution and also under successive self-multiplication.

Exercise. Graph $\cos^{10} s$ and $[J_0(s)]^{10}$ for small s.

How the remainder term behaves in general is discussed by Khinchin (1934), and obviously not all original profiles lead to the Gaussian result under continued self-convolution; for example, $\Pi(x) \sin 2\pi x$ does not. However, the functions which do manage to smooth themselves out into regular Gaussian form are numerous and include unsymmetrical functions such as $e^{-x}H(x)$ and rather unlikely looking discontinuous functions, such as $\Pi(x)$. Even $\mathrm{II}(x)$, which has two infinite discontinuities, comes close. The sometimes surprising rapidity with which the

approach to Gaussian form sets in is interesting to investigate numerically by trial. (For example, form the successive serial products $\{1 \quad 1\}^{*n}$, $\{1 \quad 1 \quad 1 \quad 1\}^{*n}$, $\{3 \quad 2 \quad 1\}^{*n}$, and test by plotting the logarithms of the terms against the square of the distance from the center.)

The central-limit theorem in its general form states that under applicable conditions the convolution of n functions (not necessarily all the same as in the discussion above) is equal to a Gaussian function whose variance is the sum of the variances of the different functions, plus a remainder which diminishes with increasing n in a certain way.

Without studying at length the conditions of applicability, we can see that the essential thing is that the Fourier transform of each of the functions entering into the convolution should exhibit the humped kind of behavior at the origin, which is possessed by our example; that is, it should be of the form $a - bs^2$. The quantity a must not be zero; that is, no one of the n functions may have zero area, or else the product of the transforms will be zero at $s = 0$, and the total convolution will be reduced to zero area. Similarly, it must not be infinite. It does not matter if b is occasionally zero, but it should in no case be infinite. Clearly, nothing resembling $\Lambda(s)$ at the origin should be included, for one such factor would leave a permanent corner at the origin, thus preventing approach to the Gaussian form.

It is not necessary that all the transforms have zero slope at the origin, but this could always be ensured by referring each function to its centroid, provided that the slope was not infinite.

The conditions may be expressed in the x domain as follows. First consider the case of identical real functions $f(x)$,

$$f(x) * f(x) * \ \dots \ * f(x),$$

and eliminate the possibility of $f(x)$ having negative area. Then the first condition ($a \neq o$ or ∞) becomes

$$0 < \int_{-\infty}^{\infty} f(x)\, dx < \infty.$$

The condition on $F'(0)$ becomes

$$\int_{-\infty}^{\infty} xf(x)\, dx = 0.$$

The requirement that $F(s)$ have a finite nonpositive second derivative (b finite) becomes

$$0 \leq \int_{-\infty}^{\infty} x^2 f(x)\, dx < \infty.$$

The possibility should also be considered that there is a discontinuity in the second derivative; but if $f(x)$ is real, the real part of $F(s)$, being even, must have the

TABLE 8.5
Table of correspondences

$\int_{-\infty}^{\infty} f(x)\,dx = F(0)$	Area equals central ordinate of transform
$\int_{-\infty}^{\infty} xf(x)\,dx = \dfrac{F'(0)}{-2\pi i}$	First moment $\propto$ central slope of transform
$\langle x \rangle = -\dfrac{F'(0)}{2\pi i F(0)}$	Abscissa of centroid $\propto$ central slope over central ordinate of transform
$\int_{-\infty}^{\infty} x^2 f(x)\,dx = -\dfrac{F''(0)}{4\pi^2}$	Second moment $\propto$ central downward curvature of transform
$\langle x^2 \rangle = -\dfrac{F''(0)}{4\pi^2 F(0)}$	Mean-square value of x $\propto$ central curvature over central ordinate of transform
$\langle x \rangle_{f*g} = \langle x \rangle_f + \langle x \rangle_g$	Abscissas of centroids add under convolution
$\langle x^2 \rangle_{f*g} = \langle x^2 \rangle_f + \langle x^2 \rangle_g$	Mean-square abscissas add under convolution
$\sigma_{f*g}^2 = \sigma_f^2 + \sigma_g^2$	Variances add under convolution

If the kth derivative of $f(x)$ becomes impulsive, $F(s) \sim |s|^{-k}$.

If the mth derivative of $f(x)$ and the nth derivative of $g(x)$ become impulsive, the $(m+n)$th derivative of $f * g$ becomes impulsive.

$W_f = \dfrac{\int_{-\infty}^{\infty} f(x)\,dx}{f(0)}$	The equivalent width of $f(x)$ is its area over its central ordinate
$W_f = \dfrac{1}{W_F}$	The equivalent width of a function is the reciprocal of the equivalent width of its transform
$\lvert f(x) \rvert \leq \int_{-\infty}^{\infty} \lvert F(s) \rvert\,ds$	Maximum possible magnitude
$\lvert f'(x) \rvert \leq 2\pi \int_{-\infty}^{\infty} \lvert sF(s) \rvert\,ds$	Maximum possible slope
$[\int f(x)g(x)\,dx]^2 \leq \int [f(x)]^2\,dx \int [g(x)]^2\,dx$	Schwarz's inequality
$\sigma_f \sigma_F \geq \dfrac{1}{4\pi}$	Uncertainty relation
$\Delta f(x) = 2\,\mathrm{I}_{\mathrm{I}}(x) * f(x)$	Finite differencing is convolution with odd impulse pair
$\Delta_a f(x) \supset 2i \sin \pi as\, F(s)$	Finite differencing over interval a corresponds, in the transform domain, to multiplication by $2i \sin \pi as$. (Note: Differentiation corresponds to multiplication by $i2\pi s$.)
$a^{-1}\Pi\left(\dfrac{x}{a}\right) * f(x) \supset \operatorname{sinc} as\, F(s)$	Taking running means over interval a corresponds to multiplication by sinc as
$\lim_{n\to\infty} f_1 * f_2 * \ldots * f_n = e^{-ax^2}$	The convolution of many functions approaches Gaussian form, provided that their transforms are humped at the origin (central-limit theorem)

same curvature to each side of the origin. Therefore any discontinuity must be confined to the imaginary part of $F''(s)$. But the imaginary part of $F(s)$ is zero at $s = 0$; therefore any effect of a discontinuity in $F''(s)$ at $s = 0$ will die out in the limit. It is therefore sufficient to require $f(x)$ to have finite area, finite mean, and finite variance.

In the event of nonidentical functions being convolved, a finite absolute third moment is required plus a more elaborate condition due to Lyapunov to ensure that the third moments are not too strong.

Exercise. Consider the behavior of $(\operatorname{sinc} x)^{*n}$, $(\operatorname{sinc}^2 x)^{*n}$, $[(1 + x^2)^{-1}]^{*n}$, $[x\Pi(x)]^{*n}$, $[\Pi(x) \sin x]^{*n}$, $[e^{-\alpha x} \sin \beta x H(x)]^{*n}$.

If one of the transforms falls to zero for some finite value $s = s_1$ the product of all the transforms will have a zero too, and so their product cannot be Gaussian. Therefore none of the convolving functions of x may be a rectangle function. This may be a moot point because the transform values of the full convolution may be so close to Gaussian for $s > s_1$ that a zero value is indistinguishable from the exact Gaussian value for some intents and purposes. This is the case with $[\Pi(x)]^{*n}$.

SUMMARY OF CORRESPONDENCES IN THE TWO DOMAINS

The results of the preceding discussion are tabulated in Table 8.5 for reference.

BIBLIOGRAPHY

Abramovitz, M., and I. A. Stegun: "Handbook of Mathematical Functions," U.S. Dept. of Commerce, Washington, D.C., 1964.

Doetsch, G.: "Theory and Application of the Laplace Transformation," Dover Publications, New York, 1943.

Jones, R. C.: "Phenomenological Discussion of the Response and Detecting Ability of Radiation Detectors," *Proc. IRE*, vol. 47, p. 1496, 1959.

Khinchin, Ya.: "Korrelationstheorie der stationären stochastischen Prozesse," *Math. Ann.*, vol. 109, pp. 604–614, 1934, reprinted in "Statistical Mechanics," Dover Publications, New York, 1949.

Uffink, J., and J. Hilgevoord: *Foundations of Physics*, vol. 15, pp. 925–945, 1985.

Van der Merwe, A., et al., eds.: "Microphysical Reality and Quantum Description," pp. 91–114, 1988.

Wiener, N.: "The Fourier Integral and Certain of Its Applications," Cambridge University Press (reprinted by Dover, New York), 1993.

PROBLEMS

1. Deduce a simple expression for $\exp(-x^2) * \exp(-x^2)$. ▷

2. Show that the self-convolution of $(1 + x^2)^{-1}$ is identical with itself, except for scale factors. Show that the self-convolution is twice as wide as the original function, that is, that the width is additive under convolution in this case, and reconcile this with the known general fact that variance is additive.

3. Investigate the functional form of $(1 + x^2/a^2)^{-1} * (1 + x^2/b^2)^{-1}$ and its width in terms of the widths of the convolved functions.

4. Show by direct integration that

$$\Lambda(x)H(x) \supset \frac{1}{i2\pi s} - \frac{e^{-i\pi s}\,\mathrm{sinc}\, s}{i2\pi s};$$

verify the result by applying first the addition and shift theorems to find the Fourier transform of $(d/dx)[\Lambda(x)H(x)]$, and then the derivative theorem in reverse.

5. By separation of the preceding transform into real and imaginary parts, show that

$$\Lambda(x)H(x) \supset \tfrac{1}{2}\,\mathrm{sinc}^2 s + \frac{1 - \cos \pi s\, \mathrm{sinc}\, s}{i2\pi s};$$

as checks on the algebra, verify that the central ordinate and slope are connected in the appropriate way with the area and first moment of $\Lambda(x)H(x)$, respectively, and note whether the transform is indeed hermitian. Break $\Lambda(x)H(x)$ into its even and odd parts, and obtain the transform of each separately.

6. Because of the irregular variation in the number of sunspots, the sequence of daily sunspot numbers is smoothed by taking five-day running totals; that is, for each day we add the sunspot number for the preceding and following two days. Here is a sequence of five-day running totals beginning Jan. 1, 1900:

45, 35, 25, 15, 5, 0, 0, 0, 0, 0, 15, 50, 80, 100, 125, 125, 100, 80, 70, 45, 30, 30, 30, 35, 60, 80, 90, 95, 100, 90, 85, 75. ▷

From this smoothed sequence, what can be deduced about the actual daily values?

7. Show that

$$W_{f*g} = \frac{W_f W_g}{W_{fg}}$$

where W_f is the equivalent width of $f(x)$.

8. Show that squares of equivalent widths are additive under convolution of Gaussian functions.

9. Show that the equivalent width of $4 \operatorname{sinc}^2 2s - \operatorname{sinc}^2 s$ is $\frac{1}{3}$, and calculate the equivalent width of $\operatorname{sinc} s + \operatorname{sinc}^2 2s$.

10. Investigate the properties of the Jones bandwidth of $f(x)$, which is defined as

$$\frac{\int_0^{\infty} FF^* \, ds}{F_{\max}}. \triangleright$$

11. Show that the autocorrelation width of $f(x)H(x)$ is twice that of $f(x)$.

12. Verify the following autocorrelation widths.

Function:	$x\Pi(x - \frac{1}{2})$	$\Pi(2x + \frac{3}{4}) + \Pi(2x - \frac{3}{4})$	$e^{-x^2/2\sigma^2}$
Width:	$\frac{3}{4}$	1	$2\pi^{\frac{1}{2}}\sigma$

13. The abscissa x of a function $f(x)$ is divided into a finite number of finite segments (plus two semi-infinite end segments). The finite segments are rearranged without overlapping, thus defining a new function which we may describe as derivable from $f(x)$ by shuffling. For example, the string of 11 pulses

$$\sum_{n=-5}^{n=+5} \Pi(11x - n)$$

and the function $|x|\Pi(x/2)$ are derivable respectively from $\Pi(x)$ and $\Lambda(x)$ by shuffling, and vice versa. Show that the equivalent width is unaffected by shuffling if the segment containing $x = 0$ is not dislodged in the shuffle, and that the autocorrelation width is not affected in any case. Consider the effect of shuffling on the total energy of a waveform, and mention several parameters of its power spectrum which are invariant under shuffling. $\triangleright$

14. An even function $g(x)$ is derived from a function $f(x)$ by the process of symmetrization illustrated in Fig. 8.21, where $AB = CD$, $EF = GH + IJ$, and so on. This process, known

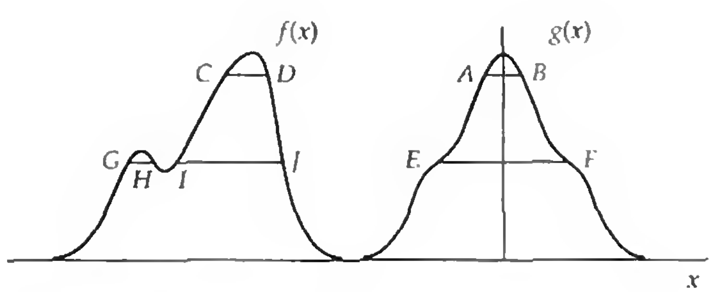


Fig. 8.21 **Steiner symmetrization.**

as Steiner symmetrization, was used by Jacob Steiner to prove that the circle is the figure of minimum perimeter for a given area. Show that the autocorrelation functions of $f(x)$ and $g(x)$ have the same equivalent width.

15. State the relationship between the functions $\exp(-x)H(x)$ and $\exp(-|x|)$ in the light of symmetrization, and say how their respective power spectra are related. ▷

16. Show that the convolution of two odd functions is even.

17. Establish the following relations between functions and their autocorrelation functions.

Function	Autocorrelation
$\Pi(x)$	$\Lambda(x)$
$e^{-\pi x^2}$	$2^{-\frac{1}{2}}e^{-\frac{1}{2}\pi x^2}$
$\delta(x)$	$\delta(x)$
$e^{-x}H(x)$	$\frac{1}{2}e^{-\|x\|}$
$e^{-\|x\|}$	$e^{-\|x\|}(1+\|x\|)$

18. Show by a simple argument in the Fourier transform domain that the autocorrelation function of $\exp(-\pi x^2)\cos\omega x$ is $2^{-3/2}\exp(-\frac{1}{2}\pi x^2)\cos\omega x$ when ω is large. ▷

19. Some difficulty arises over the autocorrelation of $\cos x$. Show that there is a sense in which the autocorrelation of a cosine function is also a cosine function.

20. Show that the product of the autocorrelation widths of a function and its transform is given by

$$W_{f\star f^*}W_{F\star F^*} = \frac{(\int f\,dx)^2(\int F\,ds)^2}{(\int f^2\,dx)^2}$$

and that the product does not have a nonzero lower limit.

21. Show that

$$|f * g| \leq \int_{-\infty}^{\infty} |FG|\,ds.$$

22. Self-reciprocal transform. We know that there is at least one function $f(\cdot)$ that is its own Fourier transform $F(\cdot)$, because

$$e^{-\pi x^2} \supset e^{-\pi s^2}.$$

It is reported that

$$f(x) = e^{-\pi(x/5)^2} + 5e^{-\pi(5x)^2}$$

is another function that is its own Fourier transform, that is, that $F(s) = f(s)$. Can you prove or disprove this claim?

23. Wavepacket spectrum. Show that

$$\exp(-\beta x^2)\cos\alpha x \supset \left(\frac{\pi}{\beta}\right)^{\frac{1}{2}} \exp\left[-\left(\frac{\alpha^2 + 4\pi^2 s^2}{4\beta}\right)\right] \cosh\left(\frac{\pi\alpha s}{\beta}\right).$$

24. Surfing. There is a tradition in surfing communities that every seventh wave is bigger. What would be the corresponding feature in the wave-spectrum domain?

25. Hermite polynomial pairs. Derive the following Fourier transform pairs

$$\begin{aligned} xe^{-\pi x^2} &\supset -ise^{-\pi s^2} \\ (4\pi x^2 - 1)e^{-\pi x^2} &\supset -(4\pi s^2 - 1)e^{-\pi s^2} \\ (4\pi x^3 - 3x)e^{-\pi x^2} &\supset i(4\pi s^3 - 3s)e^{-\pi s^2} \\ (16\pi^2 x^4 - 24\pi x^2 + 3)e^{-\pi x^2} &\supset (16\pi^2 s^4 - 24\pi s^2 + 3)e^{-\pi s^2} \end{aligned}$$

and, in general,

$$H_n(\sqrt{2\pi}\,x)e^{-\pi x^2} \supset (-i)^n H_n(\sqrt{2\pi}\,s)e^{-\pi s^2},$$

where $H_n(x)$ are the Hermite polynomials 1, $2x$, $4x^2 - 2$, $8x^3 - 12x$, $16x^4 - 48x^2 + 12$, $\ldots$, $n = 0, 1, 2, \ldots$.

26. Computed music. Describe how it might be possible to generate a sound of perpetually rising pitch that never rises beyond the range of audibility. ▷

27. Acoustic perception. A musician with his back to a violinist is able to tell whether the player is slowly moving away or is playing diminuendo. On being questioned about his ability, the musician explained, "The violinist may play a single tone loudly or softly and maintain steady tone quality, but as he moves away while sustaining a single note, there is a clear change in the color of the note. It sounds purer—the overtones less prominent." An acoustics engineer said, "When you hear a click it is followed by reverberant energy that arrives after reflection from the walls. The same click farther away sounds fainter but the amount of reverberant energy entering the ear is about the same. Therefore, the subjective impression is not the same as for a faint nearby click. Since the impulse response is different, naturally the violin note sounds different." Explain how this explanation, if it is correct, is consistent with the musician's explanation.

28. Central slope of transform phase. Let $\phi(s)$ be the phase of the Fourier transform $F(s)$ of a given real function $f(x)$. Show that, if $\phi(s)$ passes through the origin, it does so with a slope

$$\phi'(0) = -2\pi \int_{-\infty}^{\infty} xf(x)\,dx \Big/ \int_{-\infty}^{\infty} f(x)\,dx.$$

Verify the result for the case where $f(x) = xe^{-x}H(x)$.

29. Restoration for running means. Suppose that we are given $g(x)$, a function that results from smoothing $f(x)$ by convolution with a rectangle function $\Pi(x)$. Thus

$g(x) = \Pi(x) * f(x)$. We wish to find $f(x)$. See whether you can find an inverse operator $\Pi^{-1}(x)$ such that $\Pi^{-1}(x) * \Pi(x) = \delta(x)$ or whether you can find a Fourier transform for $(\operatorname{sinc} s)^{-1}$. ▷

30. Restoration for weighted running mean. Attempt to find an inverse operator for running means taken with triangular weighting. Thus if

$$g(x) = \Lambda(x) * f(x)$$

seek $\Lambda^{-1}(x)$ with the property that, if $g(x)$ is given, $f(x)$ may be found by performing the inverse operation: $f(x) = \Lambda^{-1}(x) * g(x)$. If this approach works, it will be necessary that $\Lambda^{-1}(x) * \Lambda(x) = \delta(x)$. It has been proposed that

$$f(x) = g''(x-1) + 2g''(x-2) + 3g''(x-3) + \dots.$$

Examine this formula for some special case by graphing the first two or three terms and attempt to derive it.

31. Moments. Determine constants a, b, and c such that $f(x) = a + b\cos 2\pi cx$ is a good fit to sinc x over the range $-1 < x < 1$. In what way could $F(s)$ be said to resemble $\Pi(s)$?

32. Sinc function properties. From p. 75 of Abramovitz and Stegun (1964) we discover that

$$\operatorname{sinc} x = \prod_{n=1}^{\infty}\left(1 - \frac{x^2}{n^2}\right).$$

We may think of each parabolic factor progressively forcing the product through $x = 1, 2, \dots, n$. Show that the coefficients in the expansion of $(1-x^2)(1-x^2/4)(1-x^2/9)$ do not agree with the coefficients of the Taylor series expansion of sinc x. Why is that? Why does the central limit theorem not seem to apply? ▷

33. Dual lines of reasoning. Problems that are stated in the time domain and require answers in the frequency domain can be reasoned out in either of two ways: Translate the statement into the frequency domain and find the answer to the new question, or solve in the time domain and translate the answer into the frequency domain. Consider the following problem. "A short pulse has a more or less flat spectrum up to a roll-off frequency that is some fraction of (pulse duration)$^{-1}$, but is it true that the spectrum of a waveform consisting of two identical pulses in succession, far from being approximately uniform, is such that there are frequencies where there is no content at all?"

A line of reasoning leading to this result is as follows: "The pulse pair is expressible as a convolution between the waveform of a single pulse and an impulse pair $\frac{1}{2}\delta(t) + \frac{1}{2}\delta(t-T)$, where T is the interval between pulses. Consequently, its spectrum contains a factor $\cos \pi Tf$, which produces zeros at certain frequencies." The problem did not require the full spectrum to be calculated and did not ask for the values of frequency at the zeros, but as with many important questions, was merely concerned with whether a certain phenomenon existed. Consequently, the reasoning presented can be brief. Now give the other line of reasoning leading to the same result.

34. Variance of wavepacket abscissa. Derive the variance $\langle x^2 \rangle$ of the wavepacket $f(x) = \exp[-\pi(x/W)^2]\cos 2\pi\nu x$. ▷

35. A function consisting of an asymmetrical triangular peak, with its reflection, is defined by

$$f(x) = \begin{cases} k|x|, & 0 \le |x| < a \\ 1 - |x|, & a \le |x| < 1 \\ 0, & |x| > 1 \end{cases}$$

where k and a lie between 0 and 1. Find its Fourier transform $F(s)$. ▷

36. Radar pulse generator. In order to deliver a megawatt of radio frequency power output to an antenna in a pulse of $\Delta = 0.1$ μs duration, a generator has to be excited at its input terminals by a voltage $15{,}000\Pi(t - \frac{1}{2}\Delta)$ volts. A way of doing this would be to spend a millisecond or so charging one conductor of a transmission line to a steady voltage of 30 kV. The characteristic impedance Z_0 of the transmission line would be made equal to the input resistance of the generator. At $t = 0$, a switch would connect the charged transmission line to the generator, applying 15 kV to the generator input. Electric charge would pour from the transmission line at a constant rate for 0.1 μs until the total charge was expended, whereupon the generator excitation would drop to zero and the r.f. pulse going to the transmitting antenna would terminate. (*a*) What would the length of the transmission line have to be (in meters)? (*b*) As seen from the generator, what would be the input impedance to the transmission line segment and the voltage transfer function, as a function of frequency? (*c*) What would the impulse response of the transmission line segment be? ▷

37. Functions to be expressed as finite differences. Do the exercise relating to Fig. 8.17. ▷

38. Keeping up with periodicals. Scan recent issues of journals in the current field of study, find a paper that interests you, and submit a synopsis in a form that would equip your instructor to give a 10-minute talk to the class about the work without having to refer to the original paper.

39. Composing a homework problem. Compose a homework problem suitable for your class and hand it in together with a solution suitable in final form for distribution. Bear in mind the distinction between a problem and an exercise.

40. Schrödinger's equation. A basic principle of quantum mechanics, applied to the harmonic oscillator composed of a mass m constrained by a spring of stiffness k, says that the spatial wave function $\psi(x)$ (whose squared modulus $\psi\psi^*$ gives the relative probability of finding the mass in position $x \pm \frac{1}{2}dx$), obeys the second-order, linear differential equation

$$\frac{\hbar}{2m}\frac{d^2\psi}{dx^2} + E\psi - \tfrac{1}{2}kx^2\psi = 0.$$

Solutions, for the various allowed energies $E_n = \hbar\omega(n + \frac{1}{2})$, $n = 0, 1, 2, \ldots$ are

$$\psi_n(x) = \sqrt{\alpha/\sqrt{\pi}2^n n!}H_n(\alpha x)e^{-\alpha x^2/2},$$

where $\omega^2 = k/m$ and $\alpha = (mk/\hbar)^{1/4}$. The H_n are the Hermite polynomials listed above in Problem 8.25, where the Hermite-Gauss functions $H_n(x)\exp(-x^2/2)$ can be seen to be their own Fourier transforms. That seems to imply that if one takes the Fourier transform of this Schrödinger equation term by term, the resulting differential equation for the transform Ψ must still be a Schrödinger equation. Is that so? ▷

9

Waveforms, Spectra, Filters, and Linearity

In this chapter the response of a linear system, such as an electrical filter, is considered from the point of view of resolution of the input into exponential, or harmonic, components, and an equivalent way of looking at filter action in the time domain is exhibited. The fundamental connections between linearity and convolution are also derived.

ELECTRICAL WAVEFORMS AND SPECTRA

An electrical waveform $V(t)$ is a single-valued real function of time t subject only to the general requirement of representing physically possible time dependence. Yet it is firmly established usage to speak as though steps, impulses, and absolutely monochromatic signals were electrical waveforms, even though it is clearly not possible to generate discontinuous, infinite, or eternal currents or voltages. When we make statements about the behavior of circuits in the presence of steps or impulses, we imply that a more rigorous statement exists in which that behavior would appear as the limiting form of behavior under a sequence of stimuli approaching a discontinuous step or infinite pulse. The same applies to statements about behavior in the presence of alternating or direct current.

The spectrum $S(f)$ of the electrical waveform $V(t)$ is defined as its Fourier transform (or transform in the limit as necessary):

$$S(f) = \int_{-\infty}^{\infty} V(t)e^{-i2\pi ft}\,dt$$

whence

$$V(t) = \int_{-\infty}^{\infty} S(f)e^{i2\pi ft}\,df.$$

Since $V(t)$ is by definition restricted to real functions, $S(f)$ is also subject to a restriction on its generality. The real part of a spectrum function $S(f)$ must always be even and its imaginary part odd (see Chapter 2); that is, $S(f)$ is hermitian.

Waveforms may also be subject on occasion to special restrictions of which some of the most common are given in Table 9.1.

The transform formulas for waveforms and spectra have been written in system 1 of Chapter 2. They are sometimes written in system 2 and then appear as

$$S(\omega) = \int_{-\infty}^{\infty} V(t)e^{-i\omega t}\,dt$$
$$V(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S(\omega)e^{i\omega t}\,d\omega.$$

The substitution of ω for $2\pi f$ introduces the coefficient $1/2\pi$, which, of course, may be written in front of either integral according as x and s in the system 2 formulas are identified with t and ω, or vice versa. The form given here is widely used, some users remembering the location of the unsymmetrical coefficient by the fact that if the factor $d\omega/2\pi$ were replaced by df the formulas would be symmetrical. The fact that the system 1 formulas are unchanged on generalization to more than one dimension carries no weight in circuit theory, where t has no higher-order generalization.

One also encounters the symmetrical system 3, in the form

$$S(\omega) = \frac{1}{(2\pi)^{\frac{1}{2}}}\int_{-\infty}^{\infty} V(t)e^{-i\omega t}\,dt$$
$$V(t) = \frac{1}{(2\pi)^{\frac{1}{2}}}\int_{-\infty}^{\infty} S(\omega)e^{i\omega t}\,d\omega.$$

The waveform $V(t)$ may be a train of impulses of varying amplitude. In particular, the impulses may be regularly spaced, occurring at times $t = t_i$ with strengths

Table 9.1

Corresponding restrictions

Restriction on waveform		Restriction on spectrum
Real	$\operatorname{Im} V(t) = 0$	$\operatorname{Re} S(f)$ even and $\operatorname{Im} S(f)$ odd
Switches on	$V(t) = 0 \quad t < 0$	$\operatorname{Im} S(f)$ is Hilbert transform of $\operatorname{Re} S(f)$
Nonnegative	$V(t) \geq 0$	Complicated
Finite energy	$\int_{-\infty}^{\infty} V^2\,dt = W$	$\int_{-\infty}^{\infty} SS^*\,df = W$
Finite duration	$V(t) = 0 \quad \lvert t\rvert > T$	$S(f)$ fully determined by $\text{III}(Tf)S(f)$
Band-limited	$V(t)$ fully determined by $\text{III}(Mt)V(t)$	$S(f) = 0,\ \lvert f - f_0\rvert \geq M$

a_i, where i and t_i are integers. Normal continuous-time theory will then cover signals representable by the sequence of coefficients $\{a_i\}$ and directly generate the subject matter of digital signal processing (DSP).

An interesting pair of formulas with total symmetry was proposed by Hartley (1942):

$$\psi(\omega) = \frac{1}{(2\pi)^{\frac{1}{2}}}\int_{-\infty}^{\infty} V(t)(\cos \omega t + \sin \omega t)\, dt$$
$$V(t) = \frac{1}{(2\pi)^{\frac{1}{2}}}\int_{-\infty}^{\infty} \psi(\omega)(\cos \omega t + \sin \omega t)\, d\omega.$$

It is evident that $\cos \omega t + \sin \omega t$ is a Fourier kernel, as defined in Chapter 13. The Fourier theorems have counterparts (Chapter 12) and the fast Hartley algorithm has been explored in hundreds of papers since 1983 (see *Proc. IEEE*, vol. 82, March 1994, for a special issue with bibliography).

FILTERS

We shall use the term "filter" here to denote a physical system having an input and an output. Other equivalent terms are transducer, fourpole, four-terminal network, and two-port element. Although electrical terminology will be used, the present considerations will usually apply to mechanical and acoustical transducers and to equivalent devices in other fields where vibrations or oscillations are transmitted. Of course, even within the electrical field, filters may assume a wide variety of physical embodiments, from clusters of wire coils and capacitors to intricate geometrical structures in waveguide.

When a waveform $A \cos 2\pi ft$ is fed into a linear time-invariant electrical filter, the output is also harmonic, as will be proved later, but has in general a different amplitude and phase; let it be $B \cos (2\pi ft + \phi)$. Then the filter is completely specified by a certain frequency-dependent complex quantity $T(f)$, whose amplitude is given by B/A and whose phase is ϕ; thus

$$T(f) = \frac{B}{A}e^{i\phi}.$$

We refer to $T(f)$ as the transfer factor, transfer function, system function, or frequency response of the filter.

When an input $V_1(t)$ is applied to the filter and it is desired to calculate the output $V_2(t)$, one analyzes $V_1(t)$ into its spectrum, multiplies each spectral component by the corresponding transfer factor to obtain the spectrum of $V_2(t)$, and then synthesizes $V_2(t)$ from its spectrum. Thus

$$S_2(f) = T(f)S_1(f),$$

and then
$$V_2(t) = \int_{-\infty}^{\infty} T(f)S_1(f)e^{i2\pi ft}\,df. \tag{1}$$

Since multiplication of transforms corresponds to convolution of original functions, it follows that $V_2(t)$ may be derived directly from $V_1(t)$ by smoothing with a certain function which would be characteristic of the filter, that is,

$$V_2(t) = I(t) * V_1(t). \tag{2}$$

where $I(t)$ is the Fourier transform of $T(f)$.

These two procedures are shown schematically as follows:

$$\begin{array}{rcl}
V_1(t) \text{ — transform} \rightarrow S_1(f) & \qquad & V_1(t) \\
| & & | \\
\text{multiply by} & & \text{convolve with} \\
T(f) & & I(t) \leftarrow \text{transform — } T_1(f) \\
\downarrow & & \downarrow \\
V_2(t) \leftarrow \text{transform — } S_1(f) & & V_2(t)
\end{array}$$

Combining all the quantities into one diagram with time functions on the left and transforms on the right in accordance with the convention, we can summarize the interrelations concisely as follows:

$$\begin{array}{ccc}
V_1(t) & \supset & S_1(f) \\
\text{convolution} & & \text{multiplication} \\
I(t) & \supset & T(f) \\
\downarrow & & \downarrow \\
V_2(t) & \supset & S_2(f)
\end{array}$$

The characteristic waveform $I(t)$ associated with the filter is seen from (1) to be the output obtained when $S_1(f) = 1$, that is, when $V_1(t) = \delta(t)$. It is known in communications as the "impulse response" of the filter. For many purposes it is as useful as the frequency characteristic $T(f)$ as a means of specifying a filter, and it may be much easier to obtain experimentally. Equation (2) simply breaks up $V_1(t)$ into a sequence of impulses and expresses $V_2(t)$ as a sum of the responses for each component impulse.

A third means of specifying a filter is by its "step response," that is, the output corresponding to the switching on of a constant signal $V_1(t) = H(t)$. Call this response $A(t)$. Then, regarding $V_1(t)$ as broken up into a sequence of steps $V_1'(\tau)H(t - \tau)$, we have

$$V_2(t) = A(t) * V_1'(t). \tag{3}$$

Transforming this relation and comparing with (1), we have the relation of the step-function response to the transmission characteristic of the filter:

$$S_2(f) = \overline{A}(f)i2\pi f S_1(f); \tag{4}$$

therefore $$T(f) = i2\pi f\bar{A}(f),$$

whence by a further transformation we have the relation between the step-function response and the impulse response:

$$I(t) = \frac{d}{dt}A(t).$$

Retransforming (4) with a different choice of factors, we may obtain

$$V_2(t) = A'(t) * V_1(t),$$

or formulas involving higher derivatives and integrals, including those of fractional order.

Figure 9.1 shows the relationships involving the step response $A(t)$.

The procedure with filters is summarized in the following table.

Analyze input into	Specify filter by	which has the Fourier transform
Sine and cosine waves	Transmission characteristic $T(f)$	$I(t)$
Impulses	Impulse response $I(t)$	$T(f)$
Steps	Step response $A(t)$	$T(f)/2\pi if$

The Fourier integral thus arises in filter theory in any of the following forms:

$$V(t) = \int_{-\infty}^{\infty} S(f)e^{i2\pi ft}\,df$$

$$I(t) = \int_{-\infty}^{\infty} T(f)e^{i2\pi ft}\,df$$

$$A(t) = \frac{1}{2\pi i}\int_{-\infty}^{\infty} \frac{T(f)}{f}e^{i2\pi ft}\,df,$$

and the convolution integral occurs in

$$V_2(t) = \int_{-\infty}^{\infty} I(t-\tau)V_1(\tau)\,d\tau = I(t) * V_1(t) = A'(t) * V_1(t)$$

$$V_2(t) = \int_{-\infty}^{\infty} A(t-\tau)V_1'(\tau)\,d\tau = A(t) * V_1'(t).$$

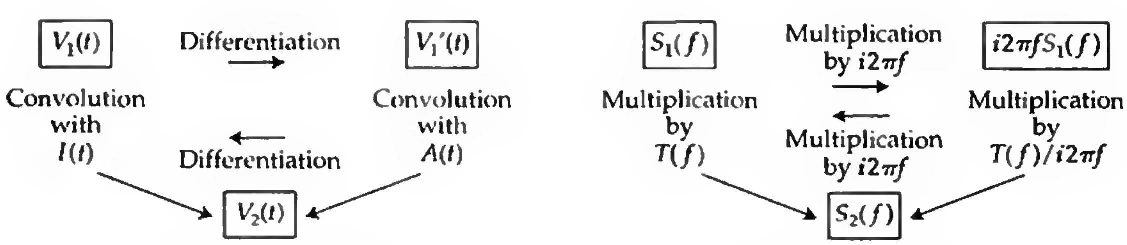


Fig. 9.1 **Operations involving the step and impulse responses, and corresponding operations in the transform domain.**

GENERALITY OF LINEAR FILTER THEORY

Throughout this section the symbol V has been used for both input and output, but the discussion is not limited to electrical quantities. Not only are mechanical, acoustical, and optical interpretations included but the input may be a voltage and the response may be a current, or vice-versa. When referring to the current response to voltage excitation, $T(f)$ has the established name transfer admittance, while transfer impedance applies to voltage response to applied current. The term *transfer factor* thus includes transfer admittance and transfer impedance as well as the dimensionless factor that applies when $V_2(t)$ and $V_1(t)$ are both voltages or both currents. Although the word "transfer," as in transfer admittance, is intended to cover cases where excitation at one point produces response at another point, nevertheless transfer factor does include response measured at the point of excitation. Transfer admittance thus includes the simple admittance at a point.

Velocity/force, flow/pressure, electromagnetic field ratios and related ratios in other subjects where linearity prevails are all covered by the foregoing section, provided time invariance (see below) also applies.

Digital signals consisting of sequences, or arrays of data such as $\{1\ 2.7\ 7.4\ 20.1\ \ldots\}$, are equivalent in information content to impulsive functions of continuous time, in this example $\delta(t) + 2.7\delta(t-1) + 7.4\delta(t-2) + 20.1\delta(t-3) + \ldots$. Thus discrete-time signals are included in the theory of this section. When it comes to numerical work, however, the practical operations are different, reducing to the convolution sums described in Chapter 3. The situation is analogous to the computation of an integral by summing discrete values, a procedure that is different from the evaluation of integrals by analysis.

Filters that are used in digital signal transmission for noise reduction, sharpening, ensuring conformity with spectral allocation, synchronizing clock circuits, and other purposes are apt to be made of electrical conductors, semiconductors, and dielectric materials configured to perform as resistors, inductances, capacitors, and active circuit elements. For the theory of networks of such elements the circuit diagram is an indispensable tool. The circuit diagram, by origin a mechanical drawing of a schematic kind showing the layout of coiled wires, spaced condenser plates, terminal posts, and connecting wires, no longer stands for a physical device at all; it stands for a set of differential equations. This is well illustrated by the migration of the circuit diagram from electrical engineering into acoustics and mechanical vibrations. The act of abstracting the diagram from the physical device requires understanding of the physics.

Once the abstraction is made, the resulting mathematical model may be handled without further reference to the physical world. Signal theory is on a further level of mathematical abstraction where the differential equations are tucked inside black boxes and replaced by overall operators. The interconnections of signal generators, filters, transmission and feedback systems, detectors, and so on, may then be represented by a flow diagram. Though clearly descended from the circuit diagram, the flow diagram suppresses circuit concepts in favor of overall

operations, such as the matrix operations that are a staple of digital signal processing theory, and is thus twice removed from manufacturing reality.

DIGITAL FILTERING

Discrete-time signals, where time t assumes integer values, can arise as regular samples of a continuous-time signal $V(t)$. The resulting sequence $f(n)$, where n is an integer, is then an approximate representation of an underlying "true" signal $V(t)$. This does not mean that $f(n)$ is inferior to $V(t)$; on the contrary, the measured values such as $f(n)$ are a basis of our knowledge about $V(t)$.

The sampling device contains a clock that connects $V(t)$ to a capacitance for a short charging time, and the voltage that builds up by the moment of disconnect is digitized and recorded. This is the sample-and-hold circuit. An apparently indirect method that is often preferable is to connect to a voltage controlled oscillator for a short time while passing the signal to a counter that counts the oscillator cycles; the count is then recorded. Whatever the device, a sample is of necessity a weighted average over a finite time.

Other discrete-time signals, such as daily counts of landings at an airport, or daily rainfall values, are not samples of any underlying continuous-time function; the discrete-time signal exists in its own right.

Discrete-time signals $f(n)$, regarded as input signals, are often subjected to filtering: low-pass filtering is applied to reduce unwanted random noise or rapid systematic fluctuations, high-pass filtering reduces unwanted drifts, and precision filtering is applied to digital signals to meet legal requirements on bandwidth that apply to telephone lines and wireless transmission.

The filtering operation is one of convolution with a desired discrete-time impulse response $h(n)$ which, in many cases, is arrived at as the Fourier transform of an imposed frequency response. Calling the output signal $g(n)$ we have the convolution sum

$$g(n) = h(n) * f(n),$$

which may be compared with the continuous-time convolution integral

$$V_2(t) = I(t) * V_1(t).$$

Chapter 3 has dealt with the machinery of discrete convolution. Chapter 13 discusses the equivalent z-transform notation that is favored in control system analysis. In theory, an impulse response $h(n)$ might have an infinite number of terms; where there is room for misunderstanding, the term "finite impulse response" (FIR) filter is available, but in numerical computing it goes without saying that impulse responses have a finite number of terms.

If a discrete-time signal $f(n)$ had to be filtered by convolution with a continuous-time impulse response $I(t)$ such as $e^{-t}H(t)$ one could readily evaluate, or graph, $V_2(t) = I(t) * [\sum f(n)\delta(t - n)]$, but the output $V_2(t)$ would no longer be a

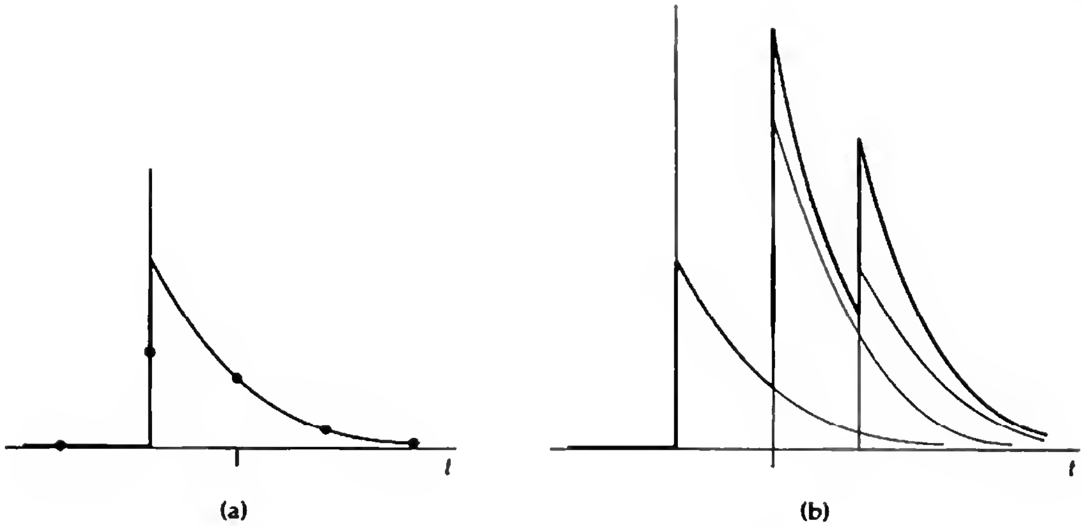


Fig. 9.2 **(a) The continuous-time impulse response $I(t) = e^{-t}H(t)$ and its samples $h(n)$ at integer values of t. (b) An input $f(n) = \{1\ 2\ 1\}$ elicits three responses, $e^{-t}H(t)$, $2e^{-(t-1)}H(t-1)$, $e^{-(t-2)}H(t-2)$, (*light curves*) that sum to $V_2(t)$ (*heavy curve*). The samples of $V_2(t)$ are identical with $g(n) = h(n) * f(n)$.**

discrete-time signal. The normal procedure would be to convert $I(t)$ into a discrete-time impulse response $h(n)$ given by $h(0) = 0.5$, $h(1) = e^{-1}$, $h(2) = e^{-2}$, ... The initial value $h(0)$ is chosen as the mean of $I(0-)$ and $I(0+)$.

Example. Given the foregoing $h(n)$ and taking $f(n) = \{1\ 2\ 1\}$,

$$\begin{aligned} g(n) &= \{0.5\ e^{-1}\ e^{-2}\ e^{-3}\ \ldots\} * \{1\ 2\ 1\} \\ &= \{0.5\ 1.37\ 1.37\ 0.69\ 0.25\ \ldots\}. \end{aligned}$$

The equivalence of a data sequence and the corresponding continuous-time impulse train is illustrated by noting that the values of $g(n)$ are the same as the samples of $V_2(t)$ at integer values of t (Fig. 9.2).

INTERPRETATION OF THEOREMS

All the theorems concerning functions and their Fourier transforms which were dealt with in Chapter 6 are interpretable in relation to waveforms, spectra, and filters. For convenience the theorems are gathered in Table 9.2 for reference, in the symbology appropriate to this chapter. The first four theorems are discussed below.

Similarity theorem. The reciprocal relationship of period and frequency is evident in this theorem, for compression of the time scale by a given factor com-

Table 9.2
Theorems for waveforms and their spectra

Theorem	Waveform $V(t)$	Spectrum $S(f)$
Similarity	$V(at)$	$\frac{1}{\lvert a\rvert}S\left(\frac{f}{a}\right)$
Addition	$V_1(t) + V_2(t)$	$S_1(f) + S_2(f)$
Shift	$V(t - T)$	$e^{-i2\pi fT}S(f)$
Modulation	$V(t)\cos \omega t$	$\frac{1}{2}S\left(f - \frac{\omega}{2\pi}\right) + \frac{1}{2}S\left(f + \frac{\omega}{2\pi}\right)$
Convolution	$I(t) * V_{in}$	$T(f)S_{in}(f)$
Autocorrelation	$V_1(t) * V_1(-t)$	$\lvert S(f)\rvert^2$
Differentiation	$V'(t)$ $V''(t)$	$i2\pi fS(f)$ $-4\pi^2 f^2S(f)$
Inverse	$tV(t)$ $t^2V(t)$	$-S'(f)/i2\pi$ $-S''(f)/4\pi^2$
Finite difference	$\Delta V(t) = V(t + \frac{1}{2}T) - V(t - \frac{1}{2}T)$	$2i \sin \pi Tf\, S(f)$
Second difference	$\Delta^2 V(t)$	$-4 \sin^2 \pi Tf\, S(f)$
Running means	$\frac{1}{T}\Pi\left(\frac{t}{T}\right) * V(t)$	$\frac{\sin \pi fT}{\pi fT}S(f)$
Rayleigh	$\int_{-\infty}^{\infty}[V(t)]^2\,dt = \int_{-\infty}^{\infty} SS^*\,df$	
Energy	$\int_{-\infty}^{\infty} V_1(t)V_2(t)\,dt = \int_{-\infty}^{\infty} S_1S_2^*\,df$	
Definite integral	$\int_{-\infty}^{\infty} V(t)\,dt = S(0)$	
Center of gravity	$\langle t\rangle = \dfrac{\int_{-\infty}^{\infty} tV(t)\,dt}{\int_{-\infty}^{\infty} V(t)\,dt} = -\dfrac{S'(0)}{2\pi i S(0)}$	
First moment	$\int_{-\infty}^{\infty} tV(t)\,dt = \dfrac{S'(0)}{-2\pi i}$	
Moment of inertia	$\int_{-\infty}^{\infty} t^2V(t)\,dt = \dfrac{S''(0)}{-4\pi^2}$	
Moment of nth order	$\int_{-\infty}^{\infty} t^nV(t)\,dt = \dfrac{S^{(n)}(0)}{(-2\pi i)^n}$	
Equivalent width	$\int_{-\infty}^{\infty} \dfrac{V(t)}{V(0)}\,dt \int_{-\infty}^{\infty} \dfrac{S(f)}{S(0)}\,df = 1$	
Inequalities	$\lvert V(t)\rvert \le \int_{-\infty}^{\infty} \lvert S(f)\rvert\,df$ $V'(t) \le 2\pi \int_{-\infty}^{\infty} \lvert fS(f)\rvert\,df$	

presses the periods of all harmonic components equally and therefore raises the frequency of every component by the same factor. Since $V(0)$ remains unaffected by time-scale changes, the area under the spectrum must, by the definite-integral theorem, remain constant, hence the compensating factor $|a|^{-1}$, which appropriately weakens the spectrum if it spreads to higher frequencies.

Addition theorem. Even when translated into the language of electrical waveforms and spectra, this theorem is simply an expression of the linearity of the Fourier transformation. It has nothing to do with the linearity of the systems in which the waveforms are found; it must, of course, be true even of waveforms in nonlinear circuits. The linear property of the transformation makes it suitable, however, for dealing with linear problems.

Shift theorem. When a waveform $V(t)$ is delayed by a given time interval T, its harmonic components are affected in different ways; for example, a component whose frequency is equal to or is an integral multiple of T^{-1} is not affected at all. Components whose period is much greater than T are not affected much, but those with periods short compared with T may be seriously altered in phase. In general, we can say that the component of period $T_1 = f_1^{-1}$ will be unchanged in amplitude but delayed in phase by $2\pi T/T_1$. Hence each component $S(f_1)$ becomes $\exp(-i2\pi f_1 T)S(f_1)$.

Modulation theorem. In the ordinary method of imposing an audio-frequency tone of angular frequency ω on a carrier wave of angular frequency Ω, the modulated waveform is

$$(1 + M \cos \omega t) \cos \Omega t,$$

where M is the depth of modulation. This waveform differs from the unmodulated carrier $\cos \Omega t$ by the addition of $M \cos \Omega t \cos \omega t$, a quantity in the form to which the modulation theorem applies.

The modulation theorem states that if a waveform $V(t)$ has a transform $S(f)$, then the waveform $V(t) \cos \omega t$ is derivable by splitting $S(f)$ into two halves, one of which is slipped to the right by an amount $\omega/2\pi$, the other going an equal distance to the left. Hence if

$$V(t) = M \cos \Omega t,$$

for which
$$S(f) = \frac{1}{2} M\delta\left(f + \frac{\Omega}{2\pi}\right) + \frac{1}{2} M\delta\left(f - \frac{\Omega}{2\pi}\right),$$

then the transform of $M \cos \Omega t \cos \omega t$ is

$$\frac{1}{4} M\delta\left(f + \frac{\Omega}{2\pi} + \frac{\omega}{2\pi}\right) + \frac{1}{4} M\delta\left(f + \frac{\Omega}{2\pi} - \frac{\omega}{2\pi}\right)$$
$$+ \frac{1}{4} M\delta\left(f - \frac{\Omega}{2\pi} + \frac{\omega}{2\pi}\right) + \frac{1}{4} M\delta\left(f - \frac{\Omega}{2\pi} - \frac{\omega}{2\pi}\right),$$

and this latter expression must be added to the spectrum of the unmodulated carrier cos Ωt (addition theorem) to give the spectrum of the amplitude-modulated wave $(1 + M \cos \omega t) \cos \Omega t$. In this way the familiar sidebands of simple modulation theory are reproduced (see Fig. 9.3).

Converse of modulation theorem. Two identical signals are sent out in succession; the spectrum of the composite signal is obtained from the spectrum of one signal alone by multiplication by a cosine function of frequency. Thus

$$V(t + T) + V(t - T) \supset 2 \cos 2\pi T f\, S(f).$$

In this expression of the converse theorem the time origin has been chosen midway between the origins of the separate signals, but one could also write (see Fig. 9.4)

$$V(t) + V(t - 2T) \supset 2e^{-i2\pi T f} \cos 2\pi T f\, S(f).$$

Both forms are derivable directly from the shift theorem.

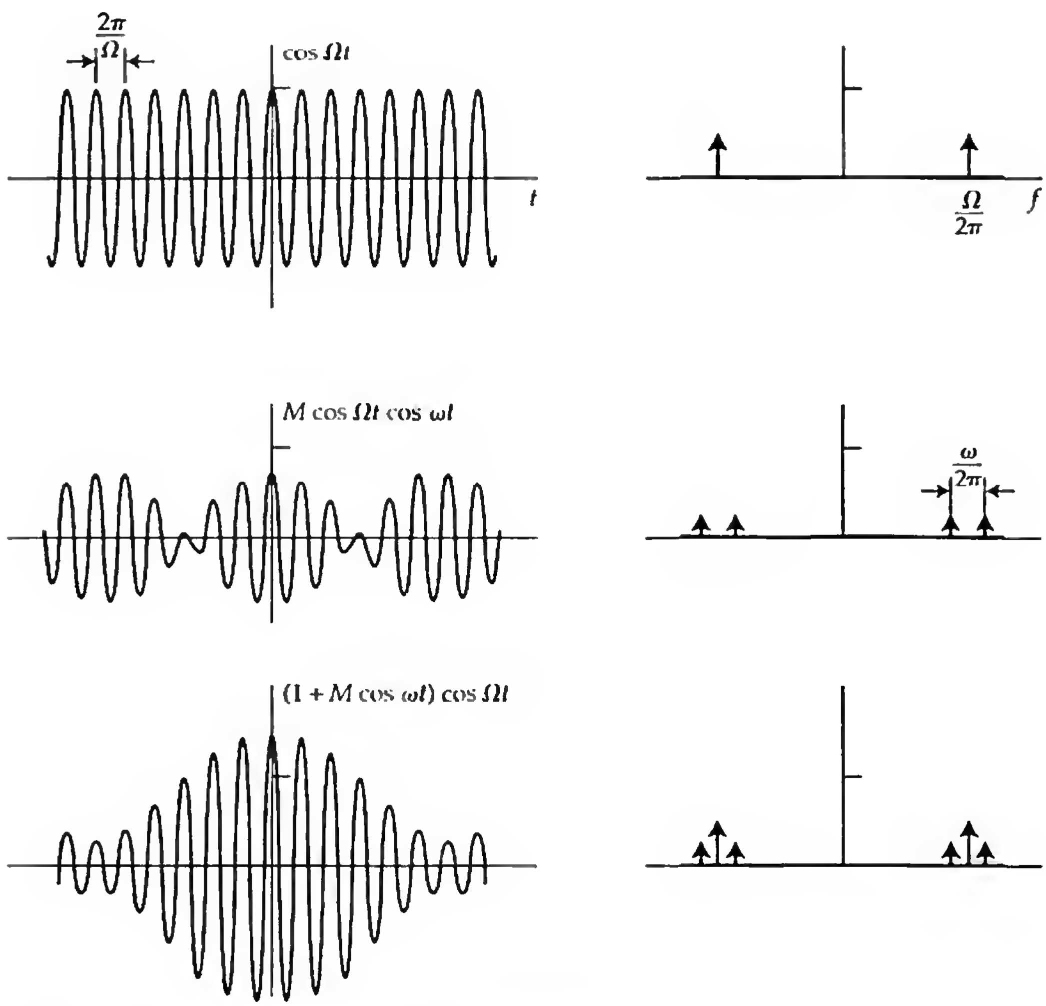


Fig. 9.3 Modulated waveforms and their spectra.

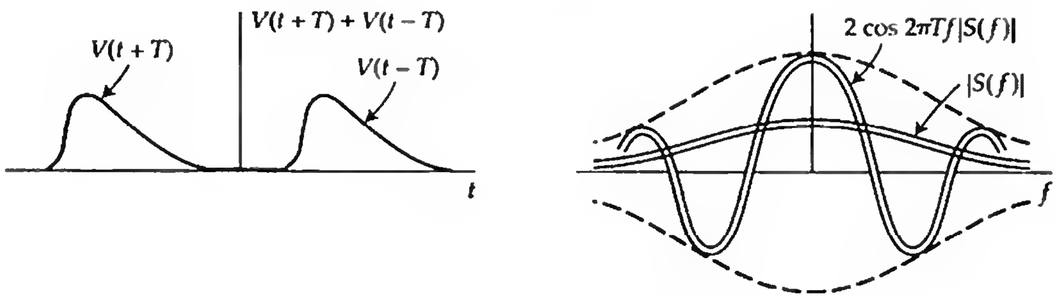


Fig. 9.4 **The converse of the modulation theorem.**

LINEARITY AND TIME INVARIANCE

Suppose that $V_2(t)$ is the response of a filter to a stimulus $V_1(t)$, and that $W_2(t)$ is the response to $W_1(t)$. Then the filter is said to be linear if the response to $V_1(t) + W_1(t)$ is $V_2(t) + W_2(t)$, irrespective of the choice of $V_1(t)$ and $W_1(t)$. Sometimes a condition is added that $aV_1(t)$ shall have a response $aV_2(t)$ for all a and $V_1(t)$, but one can deduce this relation from the superposition property (proving it first where a is an integer, then a ratio of integers[1]).

Time invariance means that the response to $V_1(t - T)$ is $V_2(t - T)$ for all T and $V_1(t)$.

As a consequence of linearity and time invariance a stimulus $A \cos 2\pi f t$, where f is any frequency, produces a response

$$B \cos (2\pi f t - \phi),$$

where ϕ and B/A may vary with frequency. Thus the response is of the same form but possibly delayed in phase by an amount ϕ and possibly different in amplitude by a factor B/A. The quantities ϕ and B/A are properties of the filter and may be compactly expressed in the form of a single complex quantity $T(f)$, the transfer factor, defined by

$$T(f) = \frac{B}{A} e^{i\phi}.$$

The essential feature of the above assertion, which will be now proved, is that a stimulus of harmonic form at a given frequency produces a response that is of the *same* form and the *same* frequency, regardless of the choice of frequency. This property is sometimes loosely referred to as "harmonic response to harmonic input."

Let the input stimulus be of unit strength and let it be the real part of a complex time-dependent function $\hat{V}_1(t)$ defined by

$$\hat{V}_1(t) = e^{i2\pi f t}.$$

---

[1] For a guide to the mathematical considerations that arise when the possibility of a voltage being irrational is not excluded *ab initio*, see Newcomb (1963).

Then the response can be shown to be given by

$$\hat{V}_2(t) = T(f)e^{i2\pi ft}.$$

The circumflex accent reminds us that, contrary to custom, we are using complex instantaneous values instead of real ones. This is done for algebraic simplicity.

To prove the assertion, let the response to $\hat{V}_1(t) = \exp(i2\pi ft)$ be $K(f,t)\exp(i2\pi ft)$, a quite general function of f and t, from which the exponential factor is extracted for convenience. Now apply a delayed stimulus $\exp[i2\pi f(t-T)]$; because of time invariance the response will be $K(f, t-T)\exp[i2\pi f(t-T)]$. However, the delayed stimulus can be expressed as the product of the original stimulus with a complex constant $\exp(-i2\pi fT)$; that is,

$$e^{i2\pi f(t-T)} = e^{-i2\pi fT}\hat{V}_1(t);$$

Hence the response can, because of linearity, be expressed as the product of the original response with this same complex factor. Thus

$$K(f, t-T)e^{i2\pi f(t-T)} = e^{-i2\pi fT}K(f,t)e^{i2\pi ft},$$

and therefore

$$K(f, t-T) = K(f,t);$$

that is, $K(f,t)$ is time-invariant and may be represented simply by $T(f)$.

The basic property of harmonic response to harmonic stimulus is thus shown to be a consequence of linearity plus time invariance.

We now show that the existence of a convolution relation

$$V_2(t) = I(t) * V_1(t)$$

between the stimulus and response is an equivalent condition.

Since the filter is linear, the response may be expressed in the form of the most general linear functional of the stimulus $V_1(t)$; that is, let

$$V_2(t) = \int_{-\infty}^{\infty} J(t,t')V_1(t')\,dt'.$$

Given also time invariance, under which $V_1(t-T)$ produces $V_2(t-T)$ for all T, it follows that

$$V_2(t-T) = \int_{-\infty}^{\infty} J(t,t')V_1(t'-T)\,dt'$$

or

$$V_2(t) = \int_{-\infty}^{\infty} J(t+T, t'+T)V_1(t')\,dt'.$$

Hence

$$J(t,t') = J(t+T, t'+T), \text{ for all } T,$$

is the condition for time invariance. It follows that $J(t,t')$ is a function purely of $t-t'$. Thus

$$J(t,t') = I(t-t')$$

and
$$V_2(t) = \int_{-\infty}^{\infty} I(t - t')V_1(t')\,dt'.$$

Hence linearity plus time invariance implies the convolution relation.

PERIODICITY

A nonconstant function $f(x)$ defined for all x is said to be periodic, with period T, if there is a positive constant T such that $f(x + T) = f(x)$, for all x. The function does not have to be continuous. A function can have more than one period, for example, $\cos x$ has a period 2π, but is also periodic with period $4\pi, 6\pi, \ldots$. The smallest period is the fundamental period. A system of impulses such as $\text{III}(x)$ does not meet the requirement of being defined for all x, but a sequence of functions defining $\text{III}(x)$ may be produced, each one of which is periodic with unit period; one could say that $\text{III}(x)$ is periodic-in-the-limit (no one does). The sum of two periodic functions is not necessarily periodic ($\cos x + \cos \pi x$) but may be ($\cos x + \cos 99x$), and may have more than one period. The presence of a periodic component does not guarantee periodicity and conversely a waveform may be periodic without any trace of a component at the corresponding frequency being detectable in the Fourier analysis (see Problems).

BIBLIOGRAPHY

Hartley, R. V. L.: "A More Symmetrical Fourier Analysis Applied to Transmission Problems," *Proc. IRE*, vol. 30, pp. 144–150, 1942.

Newcomb, R. W.: "Distributional Impulse Response Theorems," *Proc. IEEE*, vol. 51, p. 1157, 1963.

Olnejniczak, K. J., and G. T. Heydt: "Scanning the Special Section on the Hartley Transform," *Proc. IEEE*, vol. 82, pp. 372–380, 1994.

Oppenheim, A. V., and R. W. Schafer: "Discrete-time Signal Processing," Prentice-Hall, Englewood Cliffs, New Jersey, 1989.

Schwartz, M.: "Information Transmission, Modulation and Noise," 4th ed., McGraw-Hill, New York, 1990.

PROBLEMS

1. When a voltage $tH(t)$ is applied to a circuit, the response $R(t)$ is called the ramp response. Show that the response to a general applied voltage $V_1(t)$ is given by
$$\int_{-\infty}^{\infty} R(t - u)V_1''(u)\,du$$

and that this expression is particularly suited to situations where the input voltage waveform is a polygon.

2. A sound track on film is fed into a high-fidelity reproducing system at twice the correct speed. It is physically obvious, and the similarity theorem confirms that the frequency of a sinusoidal input will be doubled. Ponder what the similarity theorem says about amplitude until this is also obvious physically.

3. A transparent band (at $x = 0$) in an otherwise opaque sound track is scanned by a rectangular slit. Show that the equivalent width of the response is equal either to the width of the band or to the width of the slit.

4. A signal of finite duration is applied to a filter whose impulse response is brief in duration compared with the signal. Show that the output waveform has a longer duration than the input waveform but that the amount of stretching, as measured by equivalent width, is short compared with the duration of the impulse response.

5. The (complex) electrical length of a uniform transmission line is θ. Show that the transfer factor relating the output voltage to the input voltage is given by $T(f) = \operatorname{sech} \theta$ when no load impedance is connected.

6. A filter consists of a tee-section whose series impedances are Z_1 and Z_2 and the shunt impedance is Z_3. Show that the voltage transfer function is given by

$$T(f) = \frac{Z_3}{Z_1 + Z_3}.$$

7. In the tee-section of the previous problem, let $Z_2 = 0$ and $Z_3 = R$. The element Z_1 is an open-circuited length of loss-free transmission line of characteristic impedance R, so that we may write $Z_1 = -iR \cot 2\pi Tf$, where T is a constant. Show that the output voltage response to an input voltage step is a rectangle function of time, and hence that in general the output voltage is the finite difference of the input voltage.

8. A very large number of identical passive two-port networks are cascaded, and a voltage impulse is applied to the input, causing a disturbance to propagate down the chain. The disturbance at a distant point is expressible by repeated self-convolution of the impulse response of a single network. Does the disturbance approach Gaussian form?

9. Show that a *linear* system can be imagined whose response to a modulated signal $(1 + M \cos \omega t) \cos \Omega t$ is proportional to the audio signal $M \cos \omega t$.

10. **Fourier coefficients.** Let the y deflection of the spot on a cathode-ray oscilloscope be controlled by a periodic voltage waveform $p(t)$ of unit period so that, in some units, $y = p(t)$ and let the x deflection be controlled by a time-base generator running at a frequency n so that $x = \sin 2\pi nt$. Show that the area of the figure traced out in one period leads to a value of the Fourier coefficient a_n and that the corresponding value of b_n can be obtained similarly.

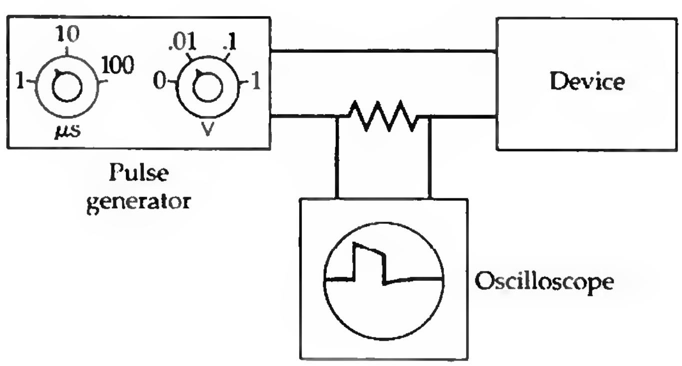
11. **Time invariance.** A two-port system contains nothing but a gate that is initially open but closes the first time a step in the input voltage occurs. Prior to this event, there is

no output, whatever the continuous input; but subsequently the output faithfully follows the input. Smith says, "It is clear to me that the response to $V_1(t - T)$ is $V_2(t - T)$ no matter what value T has and no matter what the input waveform V_1 is. Therefore, the system is time-invariant." But Jones says, "I don't agree. I fed some music into that box this morning and nothing came out, but this afternoon I fed in the same waveform and it was transmitted perfectly." Is the system (*a*) time-invariant; (*b*) linear?▷

12. **Elastic capacitor.** A parallel-plate capacitor contains solid dielectric that is squeezed when the capacitor is charged, the compression being proportional to the attractive force between the opposite charges. When a sinusoidal voltage is applied, the capacitance rises above its base value twice per cycle because the capacitance is increased when the plates are closer together. When the sinusoidal voltage is at 1000 hertz, the capacitance varies at 2000 hertz, and it is apparent that the current drawn will not be strictly sinusoidal. Consider the system whose input is the voltage and whose response is the current. Is the breakdown of sinusoidal response to sinusoidal input due to a failure of linearity, time invariance, or both? ▷

13. **Linearity and space invariance.** A glass slide carries a two-dimensional image $f(x,y)$, where the value of f is the light intensity emerging from the glass when light falls on it. A second slide $g(x,y)$ (a mask) is placed in contact with it. The quantity g is the transmission factor (ranging from 0 to 1). The effect of the mask is to multiply the original image by $g(x,y)$ to transmit a new image $h(x,y) = f(x,y)g(x,y)$.
    (*a*) Show that the operation of multiplication is linear but space variant.
    (*b*) Give an example of a particular mask to show how high spatial frequencies in the original image may be reduced to low spatial frequencies.

14. **Linearity.** A certain device, which is suspected to be nonlinear, is subjected to the following test. A pulse generator is available that can supply voltage pulses ranging in duration from 1 to 100 microseconds and with amplitudes from 1 millivolt to 1 volt, and also an oscilloscope to monitor the response current. It is observed that when the pulse duration dial is set, the waveform seen on the oscilloscope is proportional in amplitude to the voltage of the pulse. Regardless of the pulse duration selected, this proportionality is always observed. Does this establish that the device is linear? (We realize that higher voltages than those used in the test may produce nonlinear results such as sparks: therefore, we are only asking if the device is linear for amplitudes of excitation not greater than those in the tests.)



If you say the device is linear, state and prove the superposition rule that it obeys (making due provision for the allowable limits of excitation.)

If you do not admit that the device is linear, give an example of a nonlinear device that would pass the pulse generator test.

15. **Output proportional to input.** A system is such that $W_1(t)$ is the response to $V_1(t)$ and $nW_1(t)$ is the response to $nV_1(t)$, where n is integral, for all $V_1(t)$. Must the system be linear? Show that the system specification is equivalent to the definition on p. 209, or produce a system that meets the specifications and show that it is nonlinear.

16. **Associativity failure.** Two two-port networks with impulsive responses $I_1(t)$ and $I_2(t)$ are connected in tandem so that the impulse response of the composite network is $I(t) = I_1(t) * I_2(t)$. [This can be ensured to high accuracy by making the input impedance of the second network high, or can be made exact by defining $I_1(t)$ to be the impulse response at the output terminals of the first network, with the second network in place.]

    Let $I_1(t) = H(t)$, $I_2(t) = \delta(t) - \delta(t-1)$, and $I(t) = \Pi(t - \frac{1}{2})$. If the voltage $V_{\text{in}}(t)$ at the input is given by $V_{\text{in}}(t) = e^{-t}$, then the output voltage is given by $V_{\text{out}} = V_{\text{in}}(t) * I(t) = \Pi(t - \frac{1}{2}) * e^{-t} = \int_{t-1}^{t} e^{-u}\,du - (e-1)e^{-t}$. In effect, we have said that

$$V_{\text{out}}(t) = V_{\text{in}}(t) * [I_1(t)] * [I_2(t)].$$

By the associative property of convolution, we expect to be able to say that

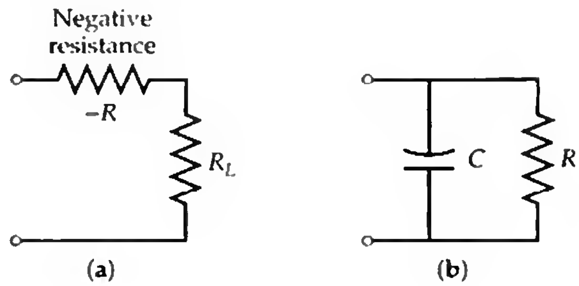
$$V_{\text{out}}(t) = [V_{\text{in}}(t) * I_1(t)] * I_2(t),$$

where the quantity in brackets is the response at the junction between the two networks. But $V_{\text{in}}(t) * I_1(t) = e^{-t} * H(t) = \int_{-\infty}^{\infty} e^{-u} H(t-u)\,du = \int_{-\infty}^{t} e^{u}\,du$, which is not a convergent integral. Discuss this breakdown of associatively.

17. **Associativity.** What can be said about the associativity of convolution in the following cases?

$$H(x) * \delta'(x) * H(-x)$$
$$H(x) * \delta'(x) * \operatorname{sgn} x$$
$$e^{-x} * [\delta(x) - \delta(x-1)] * H(x).$$

18. **Stability.** In mechanics a system is said to be stable if a small perturbation produces forces tending to return the system to its original configuration. Similar usage is found in electronics in connection with amplifiers and feedback circuits. However, a different idea is also in use according to which a linear system is said to be stable if $|V_2(t)| < kM$ for any $|V_1(t)| < M$. In other words, if the input $V_1(t)$ keeps below a maximum value M in absolute value, then, for a stable system, the output will also remain below some maximum kM for any choice of the input waveform. In another form of the definition, the system is said to be stable if the absolute integral of the impulse response, $\int_{-\infty}^{\infty} |I(t)|\,dt$, has a finite value k. Are the following systems stable? [Let $V_1(t)$ be the applied voltage and $V_2(t)$ be the current that flows.]



19. Stability. A short-circuited loss-free transmission line of length l on which the wave velocity is v has a current impulse response

$$I(t) = Z_0^{-1}\left[\delta(t) + 2\delta\left(t - \frac{2l}{v}\right) + 2\delta\left(t - \frac{4l}{v}\right) + \ldots\right].$$

Because $\int_{-\infty}^{\infty} |I(t)|\,dt$ is not finite, it seems to be unstable. On second thoughts, if a bounded input voltage $\Pi(t)$ is applied, the current response is $Z_0^{-1}[\Pi(t) + 2\Pi(t - 2l/v) + 2\Pi(t - 4l/v) + \ldots]$, which is a bounded output. Since bounded input seems to produce bounded output, it now seems to be stable. What is the truth?

20. Linearity: continuity of transformation. A mathematical system model has the following relations between input functions $V_1(t)$ and output functions $V_2(t)$; namely, $V_2(t)$ is the same as $V_1(t)$ except that any discontinuities are removed. To be precise, if $V_1(t)$ has a finite number N of discontinuities of size $J_1, J_2, \ldots, J_n$ at $t = a_1, a_2, \ldots, a_N$, then

$$V_2(t) = V_1(t) - \sum_{n=1}^{N} J_n H(t - a_n).$$

Find out whether this system is linear and time-invariant.

21. Continuity of transformation. Let $M(x)$ be a nonzero, real-valued function such as $\exp[-(1 - x^2)^{-1}]\Pi(x/2)$ that possesses (nonimpulsive) derivatives of all orders and is zero for all t outside the interval $(-1,1)$. Consider the sequence $p_\tau(x) = \tau^{-1}M[\tau^{-1}(x - \tau)]$ generated as $\tau \to 0$. The area under $p_\tau(x)$ remains constant independent of τ. Given any value of x,

$$\lim_{\tau\to 0} p_\tau(x) = 0 \qquad \text{for all } x.$$

The reason is that all the profiles $p_\tau(x)$ are identically zero for $x \le 0$; and for any positive x there is a value $\tau = x/2$ such that $p_\tau(x) = 0$ for all smaller values of τ.

Apply $p_\tau(x)$ as an input to a linear system which is represented by a transformation $\mathcal{S}$ that maps the input function $p_\tau(x)$ into an output function $q_\tau(x)$. Represent this by

$$q_\tau = \mathcal{S}[p_\tau].$$

Let the transformation $\mathcal{S}$ be continuous:

$$\mathcal{S}[\lim_{\tau\to 0} p_\tau] = \lim_{\tau\to 0} \mathcal{S}[p_\tau].$$

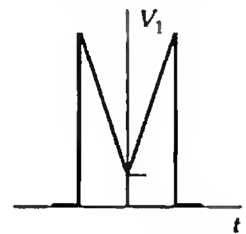
Because of linearity the LHS $\mathcal{S}[0] = 0$; therefore, the RHS

$$\lim_{\tau\to 0} \mathcal{S}[p_\tau] = 0.$$

But we would expect to elicit the impulse response on the RHS. Does this mean that ordinary physical systems are not continuous?

22. **Impulse response.** A linear time-invariant system having an impulse response $I(t)$ is excited by an input voltage $V_1(t)$, where

$$V_1(t,\tau) = [1 + \tau^{-2}(1 - 2\tau)|x|]\Pi\left(\frac{x}{2\tau}\right)$$



and produces a response $V_2(t,\tau)$.

(*a*) Is it true that $\lim_{\tau\to 0} V_2(t,\tau) = I(t)$?

(*b*) Does $\lim_{\tau\to 0} V_1(t,\tau)$ exist?

23. **Sinc function properties.** Verify that

$$f_4(x) = \cos\left(\frac{\pi x}{2}\right)\cos\left(\frac{\pi x}{4}\right)\cos\left(\frac{\pi x}{8}\right)\cos\left(\frac{\pi x}{16}\right)$$

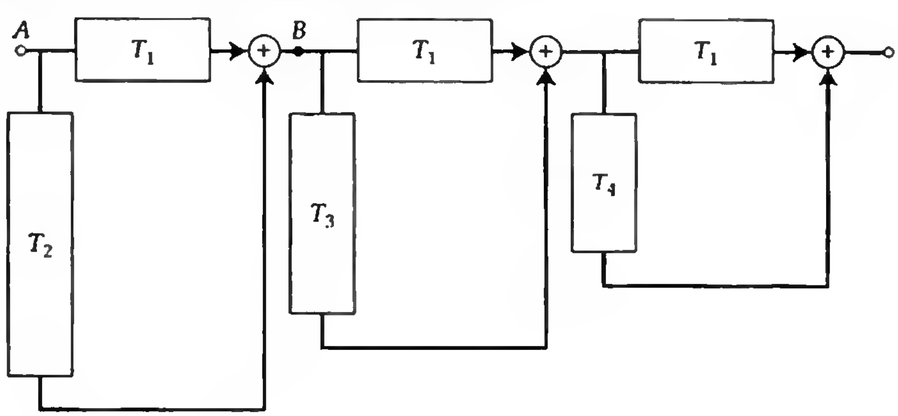
is a good approximation to sinc x by finding out how large x has to be before the error reaches 1 percent of the peak value. If the expression quoted is a good approximation to sinc x, should not its Fourier transform be a good approximation to $\Pi(s)$? Discuss this question by first working out the transform. Now show that

$$\operatorname{sinc} x = \prod_{n=1}^{\infty} \cos\left(\frac{\pi x}{2^n}\right).$$

If the continued product is taken to N factors only, instead of to infinity, show that the formula can be saved by including a rather broad correction factor sinc $(x/2^N)$:

$$\operatorname{sinc} x = \prod_{n=1}^{N} \cos\left(\frac{\pi x}{2^n}\right)\operatorname{sinc}\left(\frac{x}{2^N}\right).$$

24. **Rectangular impulse response.** A signal entering a certain system with input at A and output at B is divided into two channels, one of which contains a delay T_1 and the other a longer delay T_2, and then the two channels are brought together and added to form the output signal. Let the transfer function be $T(f)$.



(*a*) Show that T_1 and T_2 may be chosen to make $|T(f)|$ any desired cosine function of frequency.

(*b*) Show that a chain of three such systems is suitable for practical equipment that is to generate a rectangular impulse response. Explain how to choose T_3 and T_4.

25. Prolate spheroidal wavefunctions. Radial-meridional (nonlongitudinal) oscillations of electric field u around a conducting football obey the differential equation

$$(1 - t^2)\frac{d^2u}{dt^2} - 2t\frac{du}{dt} + (\chi - \omega^2 t^2)u = 0,$$

where t (later to be reinterpreted as time) is the meridional coordinate. The parameter ω is $(2\pi/\text{free-space wavelength}) \times$ semi-interfocal distance, and the constant χ is a separation constant. For each choice of χ and ω, there is a series of resonant modes $n = 0, 1, 2, \ldots$. The corresponding electric charge solutions of the differential equation, which are known as prolate spheroidal wavefunctions $S_{0n}(\omega,t)$, turn out to be of interest where a signal must be squeezed into the shortest time and simultaneously into the narrowest band. A property of the prolate spheroidal wavefunctions is that

$$\int_{1}^{1} \operatorname{sinc}\left[2\left(\frac{\omega}{2\pi}\right)(t - t')\right] S_{0n}(\omega,\omega')\, dt' = \lambda S_{0n}(\omega,t),$$

where λ is an eigenvalue. The subscript 0 refers to the absence of longitude dependence. Interpreting t as time, we may say that the integral operation consists of gating or truncating the operand to eliminate portions outside $t = \pm 1$ s, followed by sinc function smoothing that eliminates frequencies $\omega/2\pi$ beyond ± 1 Hz. The consequence of this squeezing operation is to leave $S_{0n}(\omega,t)$ unchanged except for a factor λ. The prolate spheroidal wavefunctions are thus eigenfunctions of the integral operator. The eigenvalue λ, which may be looked up in tables (Abramovitz and Stegun, p. 753), is $\lambda = [R_{0n}^{(1)}(\omega,1)]^2$.

Show that the prolate spheroidal wavefunctions are also eigenfunctions of the finite Fourier transform, that is, that

$$\int_{-1}^{1} e^{-i\omega t t'} S_{0n}(\omega,t')\, dt' = \mu S_{0n}(\omega,t).$$

26. Multiplication by x. Show that $xf(x) \supset (i/2\pi)F'(s)$.▷

27. Elusive period. (*a*) Make a plot of $y(x) = 9 \cos 5x + 11 \cos 4x$ for $0 \leq x \leq 6$, but suppose that x actually runs from $-\infty$ to ∞. Examine your sample of the function $y(x)$ and decide whether $y(x)$ meets the condition for a periodic function. If yes, what is the period? If no, explain. (*b*) Plot the Fourier transform of $y(x)$.▷

28. Periodic barcode. A pattern consisting of a set of thin black lines on white is made by printing 31 lines spaced by 5 mm of white and then overprinting with 37 lines spaced 4 mm. The first black line of the second set falls exactly on the first black line of the first set. The lines have a width $w = 1$ mm; thus each line is describable by $\Pi(x/w)$ suitably shifted. The final printing is describable by $y(x)$, where $y(x) = 0$ for white and 1 for black. The printed barcode appears periodic to the eye, but is the corresponding frequency present in the Fourier transform $Y(s)$?▷

29. **Spectral resolution.** A spectrograph responds to a spectral line $\delta(x)$ as $\exp(-\pi x^2/W^2)$. If there are two equal spectral lines spaced X apart, the instrumental response is $\exp[-\pi(x+\frac{1}{2}L)^2/W^2] + \exp[-\pi(x-\frac{1}{2}L)^2/W^2]$. If the lines are well spaced, $L >> W$, then two separate peaks are recorded with a minimum between. (*a*) As the two lines are brought closer together there is a value $L = L_{crit}$, where the central minimum rises to the level of the two peaks. What is the value of L_{crit}? (*b*) If the lines are more closely spaced, say $L = 0.8L_{crit}$, only one peak will be recorded. Would it then be possible to deduce that two lines had been involved and to determine their line spacing? ▷

30. **Filtering a digital signal.** A discrete-time signal {1 1.6 2 0.6} whose elements are spaced one microsecond is applied to a resistance-capacitance filter whose impulse response is $\exp(-t/RC)H(t)$, where $R = 1$ ohm and $C = 1$ microfarad and the time constant RC is 1 microsecond. Represent the signal by the continuous-time impulse train $V_1(t) = \delta(t-1) + 1.6\delta(t-2) + 2\delta(t-3) + 0.6\delta(t-4)$ and graph the response $V_2(t)$ as the sum of four exponentials. (*a*) Mark the values of $V_2(t)$ for $t = 0, 1, 2, 3, 4, 5, 6$ and give algebraic expressions for these values. (*b*) Compare the coefficients in these expressions with those obtained from numerical evaluation of the convolution

$$\{0.5\ e^{-1}\ e^{-2}\ e^{-3}\ e^{-4} \ldots\} * \{1\ 1.6\ 2\ 0.6\}. \triangleright$$

10

Sampling and Series

Suppose that we are presented with a function whose values were chosen arbitrarily, and suppose that no connection exists between the neighboring values chosen for the dependent variable. Thus if the independent variable were time, we would have to expect jumps of arbitrary magnitude and sign from one instant to the next.

In nature such a function would never be observed. Because of limitations of the measuring instrument, or for other reasons, there is always a limit to the rate of change. One might thus surmise that there is at least a little interdependence between waveform values at neighboring instants, and consequently that it might be possible to predict from past values over a certain brief time interval. This suggests that it might be possible to dispense with the values of a function for intervals of the same order and yet preserve essentially all the information by noting a set of values spaced at fine, but not infinitesimal, intervals. From this set of samples it would seem reasonable that the intervening values could be recovered if only to some degree of approximation.

SAMPLING THEOREM

The sampling theorem states that, under a certain condition, it is in fact possible to recover with full accuracy the values intervening between regularly spaced samples; in other words, the sample set can be fully equivalent to the complete set of function values. The condition is that the function should be "bandlimited"; that is, it should have a Fourier transform that is nonzero over a finite range of the transform variable and zero elsewhere. Digital signal processing can be applied to the sample set while at the same time the theory of functions of a continuous variable can be advantageously applied to derive and clarify the basics of DSP.

Clearly, the interval between samples is crucial in deciding the utility of the theorem; if the samples had to be very close together, not much would be gained. As an illustration of the fineness of sampling we take a simple band-limited function, namely sinc x, whose spectrum is flat where $|s| < \frac{1}{2}$ and zero beyond. The sampling interval, deduced as explained below, is 1. Figure 10.1*a* shows the sample values that define sinc x, and it will be seen that the interval is extremely coarse in comparison, for example, with the interval that would intuitively be chosen for numerical integration. The sampling intervals indicated by this theorem often seem, at first, to be surprisingly wide. Also shown in Fig. 10.1*b* are a set of samples for $\text{sinc}^2 \frac{1}{2}x$, a function whose spectrum cuts off at the same points ($|s| = \frac{1}{2}$) as that of sinc x. Figure 10.1*c* and 10.1*d* provide some samples for experiment.

With any given waveform, there is always a frequency beyond which spectral contributions are negligible for some purposes. However, on the other hand, the transform probably never cuts off absolutely; consequently, in applications of the sampling theorem, the error incurred by taking a given waveform to be band-limited must always be estimated.

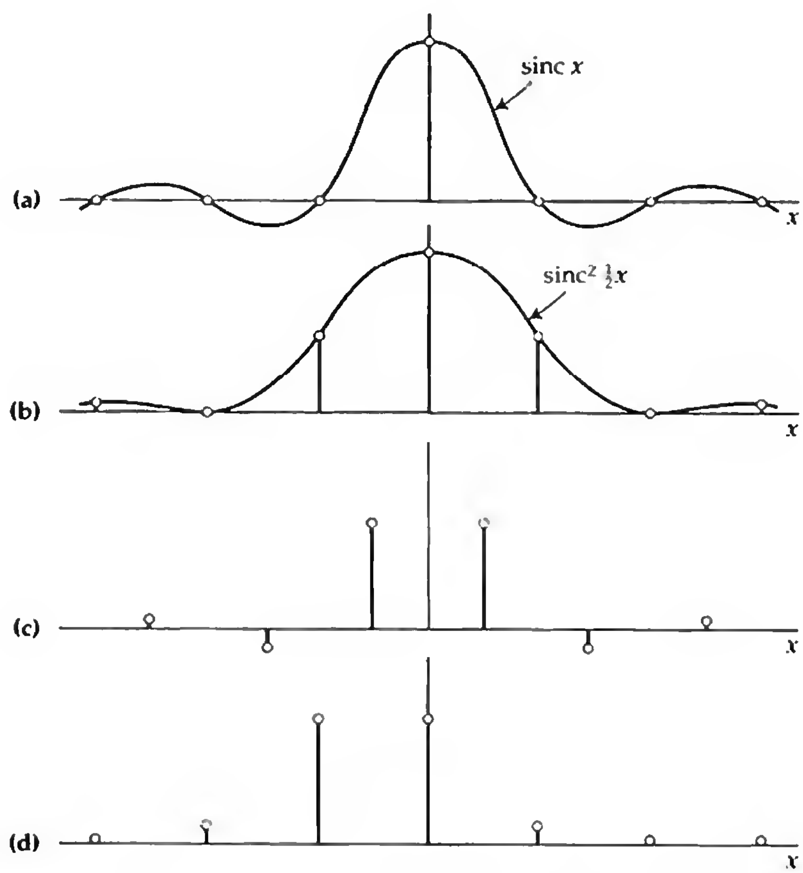


Fig. 10.1 Two functions and their samples, and two sets of samples for practice.

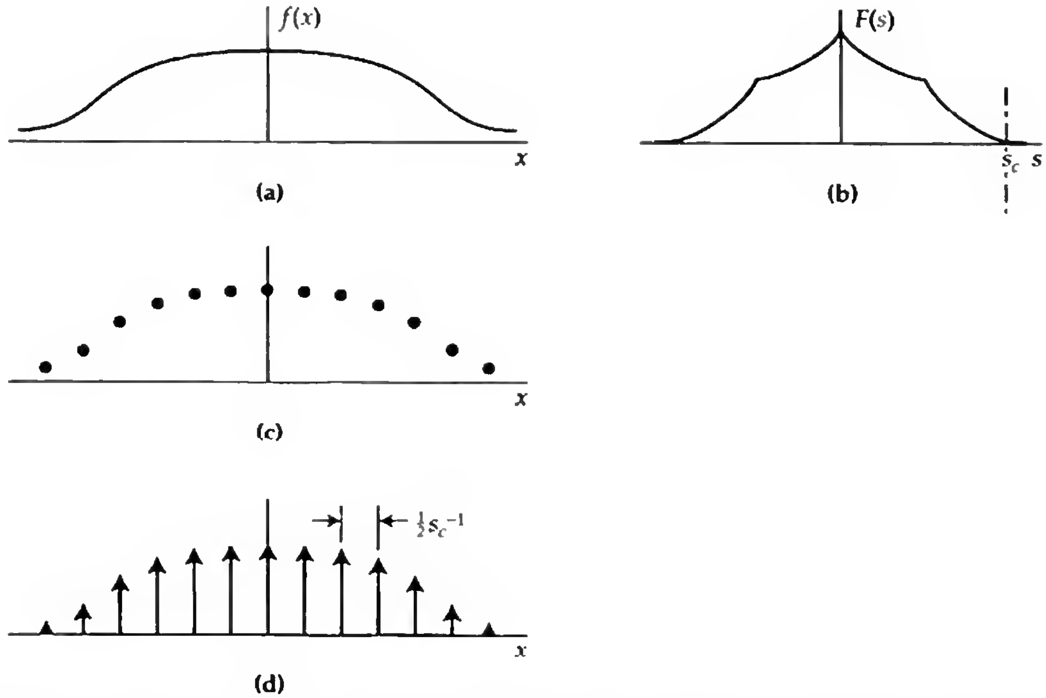


Fig. 10.2 **(a) A band-limited function $f(x)$ with a cutoff spectrum (b). The samples (c) suffice to reconstitute $f(x)$ accurately in full detail and are equivalent in content to the train of impulses (d).**

Consider a function $f(x)$, whose Fourier transform $F(s)$ is zero where $|s| > s_c$ (see Fig. 10.2). Evidently $f(x)$ is a band-limited function; in this case the band to which the Fourier components are limited is centered on the origin of s. Such a function is representative of a wide class of physical distributions which have been observed with equipment of limited resolving power. We shall refer to such transforms as "cutoff transforms" and describe them as being cut off beyond the "cutoff frequency" s_c.

In general a cutoff transform is of the form $\Pi(s/2s_c)G(s)$, where $G(s)$ is arbitrary, and therefore the general form of functions whose transforms are cut off is

$$\text{sinc } 2s_c x * g(x),$$

where $g(x)$ is arbitrary.

Of course, if the original function is cut off, then it is not a band-limited function; it is the transform which is band-limited.

With the exception noted below, band-limited functions have the peculiar property that they are fully specified by values spaced at equal intervals not exceeding $\frac{1}{2}s_c^{-1}$ (see Fig. 10.2*c*).

In the derivation that follows, the introduction of the *shah* symbol proves convenient, because multiplication by $\text{III}(x)$ is equivalent to sampling, in the sense that information is retained at the sampling points and abandoned in between.

As an additional bonus, the *shah* symbol, as a result of its replicating property under convolution, enables us to express compactly the kind of repetitive spectrum that arises in sampling theory.

In the course of the argument we use the relation

$$\mathrm{III}(x) \supset \mathrm{III}(s),$$

which is discussed at greater length at the end of this chapter.

Consider the function

$$\tau^{-1} f(x)\mathrm{III}\left(\frac{x}{\tau}\right) = \sum_n f(n\tau)\,\delta(x - n\tau)$$

shown in Fig. 10.2*c*. Information about $f(x)$ is conserved only at the sampling points where x is an integral multiple of the sampling interval τ. The intermedi-

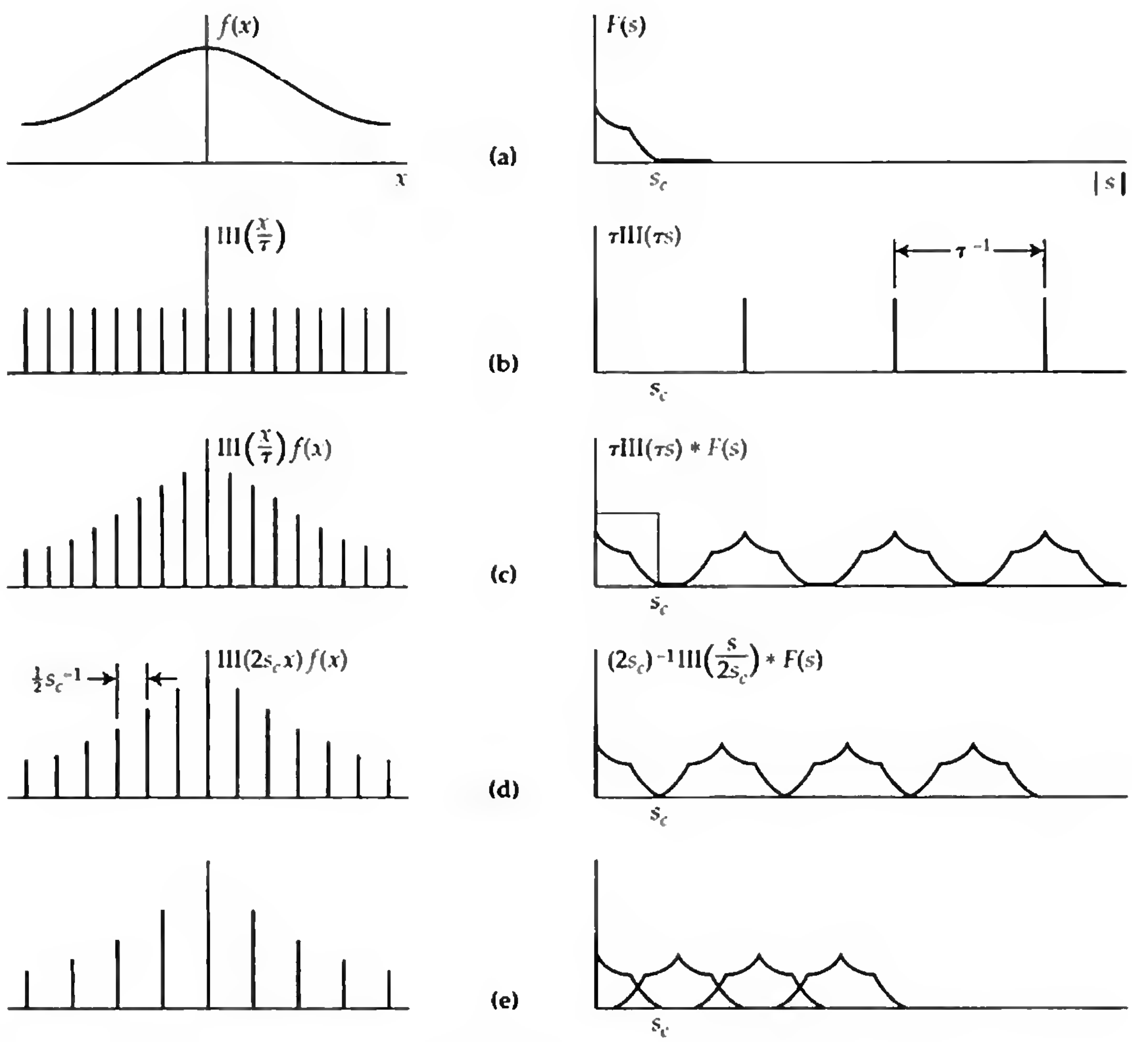


Fig. 10.3 **Demonstrating the sampling theorem.**

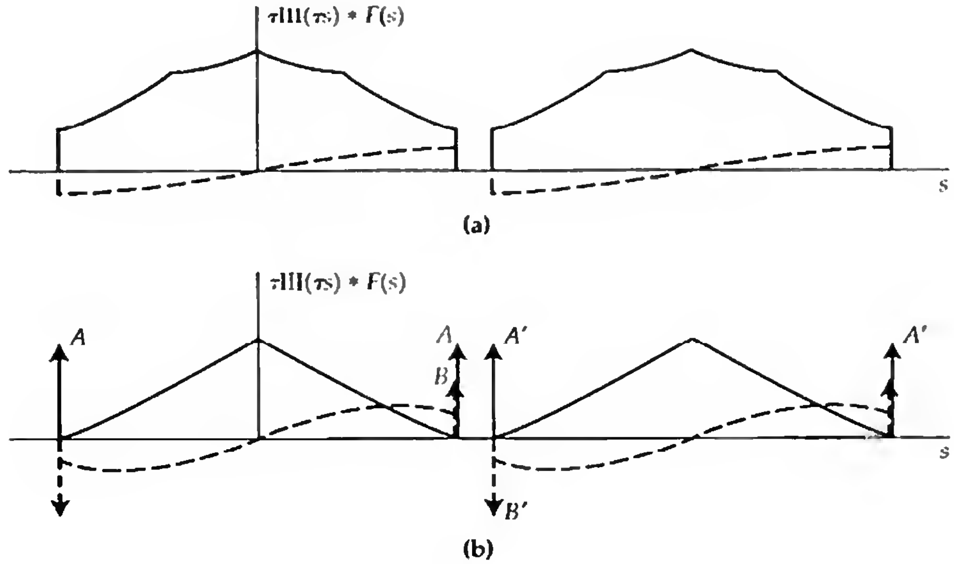


Fig. 10.4 **Critical sampling.**

ate values of $f(x)$ are lost. Therefore, if $f(x)$ can be reconstructed from $f(x)III(x/\tau)$, the theorem is proved. The transform of $III(x/\tau)$ is $\tau III(\tau s)$, Fig. 10.3*b*, which is a row of unit impulses at spacing τ^{-1}. Therefore

$$\overline{f(x)III(x/\tau)} = \tau III(\tau s) * F(s),$$

and we see that multiplication of the original function by $III(x/\tau)$ has the effect of replicating the spectrum $F(s)$ at intervals τ^{-1} (see Fig. 10.3*c*). We can reconstruct $f(x)$ if we can recover $F(s)$, and this can evidently be done, in the case illustrated in Fig. 10.3*c*, by multiplying $\tau III(\tau s) * F(s)$ by $\Pi(s/2s_c)$. Except for cases of singular behavior at $s = s_c$ (to be considered below), this is sufficient to demonstrate the sampling theorem.

At the same time a condition for sufficiently close sampling becomes apparent, for recovery will be impossible if the replicated islands overlap as shown in Fig. 10.3*e*, and this will happen if the spacing of the islands τ^{-1} becomes less than the width of an island $2s_c$. Hence the sampling interval τ must not exceed $\frac{1}{2}s_c^{-1}$, the semiperiod of a sinusoid of frequency s_c, and for critical sampling we shall have the islands just touching, as in Fig. 10.3*d*.

A small refinement must now be considered before the sampling theorem can be enunciated with strictness. In the illustration $F(s_c)$ is shown equal to zero. If $F(s_c)$ is not zero, the islands have cliffs which, under conditions of sampling at precisely the critical interval, make butt contact. In Fig. 10.4*a* this is on the point of happening, but careful examination, taking into account also the imaginary part of $F(s)$, reveals that multiplication by $\Pi(s/2s_c)$ permits exact recovery of $F(s)$. However, if $F(s)$ behaves impulsively at $s = s_c$, that is, if $f(x)$ contains a harmonic component of frequency s_c, then there is more to be said.

In Fig. 10.4*b*, which shows such a case, consider first that $F(s)$ is even; that is, ignore the odd imaginary part shown dotted. Then as the sampling interval ap-

proaches the critical value, the impulses A and A' tend to fuse at $s = s_c$ into a single impulse of double strength, and multiplication by $\Pi(s/2s_c)$ taken equal to $\frac{1}{2}$ at $s = s_c$, restores the impulse to its proper value, thus permitting exact recovery of $F(s)$. The impulses A represent, of course, an even, or cosinusoidal, harmonic component: $2A \cos 2\pi s_c x$. Now consider the impulses contained in the odd part of $F(s)$. Under critical sampling, B and B' fuse and cancel. Any odd harmonic component proportional to $\sin 2\pi s_c x$ therefore disappears in the sampling process.

Exercise. The harmonic function $\cos(\omega t - \phi)$ is sampled at its critical interval (the semiperiod π/ω). Split the function into its even and odd parts and note that the sample values are precisely those of the even part alone. Note that the odd part is sampled at its zeros.

The sampling theorem can now be enunciated for reference:

> *A function whose Fourier transform is zero for $|s| > s_c$ is fully specified by values spaced at equal intervals not exceeding $\frac{1}{2}s_c^{-1}$ save for any harmonic term with zeros at the sampling points.*

In this statement of the sampling theorem there is no indication of how the function is to be reconstituted from its samples, but from the argument given in support of the theorem it is clear that it is possible to reconstruct the function from the train of impulses equivalent to the set of samples, using some process of filtration. This procedure, which is envisaged as filtering in the transform domain, evidently amounts in the function domain to interpolation.

INTERPOLATION

The numerical process of calculating intermediate points from samples does not of course depend on calculating Fourier transforms. Since the process of recovery was to multiply a transform by $\Pi(s/2s_c)$, the equivalent operation in the function domain, namely, convolution with $2s_c \operatorname{sinc} 2s_c x$, will yield $f(x)$ directly from $\text{III}(x/\tau)f(x)$. Convolution with a function consisting of a row of impulses is an attractive operation numerically because the convolution integral reduces exactly to a summation (serial product).

For midpoint interpolation we can permanently record a table (see Table 10.1) of suitably spaced values of sinc x, and it proves practical when further interpolation is required to repeat the midpoint process, using the same array.

RECTANGULAR FILTERING IN FREQUENCY DOMAIN

Suppose that it is required to remove from a function spectral components whose frequencies exceed a certain limit, that is, to multiply the transform by a rectan-

■ TABLE 10.1
Midpoint interpolation

$\|x\|$	sinc x	$\|x\|$	sinc x	$\|x\|$	sinc x	$\|x\|$	sinc x
$\frac{1}{2}$	0.6366	$9\frac{1}{2}$	−0.0335	$18\frac{1}{2}$	0.0172	$27\frac{1}{2}$	−0.0116
$1\frac{1}{2}$	−0.2122	$10\frac{1}{2}$	0.0303	$19\frac{1}{2}$	−0.0163	$28\frac{1}{2}$	0.0112
$2\frac{1}{2}$	0.1273	$11\frac{1}{2}$	−0.0277	$20\frac{1}{2}$	0.0155	$29\frac{1}{2}$	−0.0108
$3\frac{1}{2}$	−0.0909	$12\frac{1}{2}$	0.0255	$21\frac{1}{2}$	−0.0148	$30\frac{1}{2}$	0.0104
$4\frac{1}{2}$	0.0707	$13\frac{1}{2}$	−0.0236	$22\frac{1}{2}$	0.0141	$31\frac{1}{2}$	−0.0101
$5\frac{1}{2}$	−0.0579	$14\frac{1}{2}$	0.0220	$23\frac{1}{2}$	−0.0135	$32\frac{1}{2}$	0.0098
$6\frac{1}{2}$	0.0490	$15\frac{1}{2}$	−0.0205	$24\frac{1}{2}$	0.0130	$33\frac{1}{2}$	−0.0095
$7\frac{1}{2}$	−0.0424	$16\frac{1}{2}$	0.0193	$25\frac{1}{2}$	−0.0125	$34\frac{1}{2}$	0.0092
$8\frac{1}{2}$	0.0374	$17\frac{1}{2}$	−0.0182	$26\frac{1}{2}$	0.0120	$35\frac{1}{2}$	−0.0090

gle function, which we shall take to be $\Pi(s)$. We are assuming that $s_c = \frac{1}{2}$ and that the critical sampling interval is 1. This is just the operation which has already been carried out for the purpose of interpolation. However, in general the function to be filtered will not consist solely of impulses, and the convolution integral giving one filtered value does not reduce exactly to a summation. However, when it is evaluated numerically it will have to be approximated by a summation,

$$\Sigma_\tau = f(x) * \left[\mathrm{III}\left(\frac{x}{\tau}\right)\operatorname{sinc} x\right],$$

and we may ask how coarse the tabulation interval may be and still sufficiently approximate the desired integral

$$f(x) * \operatorname{sinc} x.$$

Beginning with $\tau = 1$, we find $\Sigma_1 = f(x)$; that is, no filtering at all has resulted. Now trying $\tau = \frac{1}{2}$, we have

$$\Sigma_{\frac{1}{2}} = f(x) * \mathrm{III}(2x)\operatorname{sinc} x$$

and

$$\overline{\Sigma}_{\frac{1}{2}} = F(s)[\tfrac{1}{2}\mathrm{III}(\tfrac{1}{2}s) * \Pi(s)],$$

since

$$\overline{\Sigma}_\tau = F(s)[\tau\mathrm{III}(\tau s) * \Pi(s)].$$

Hence $\overline{\Sigma}_{\frac{1}{2}}$ consists of a central part $F(s)\Pi(s)$ plus remoter parts. For many purposes this simple operation would suffice (for example, when the components to be rejected lie chiefly just beyond the central region).

Adopting an idea from the method of interpolation, where we economize on interpolating arrays by repeated use of the one midpoint interpolation array, we now consider the effect of repeated approximate filtering of the one kind. By using the same filtering array at $\tau = \frac{1}{4}$ we have the filter characteristic

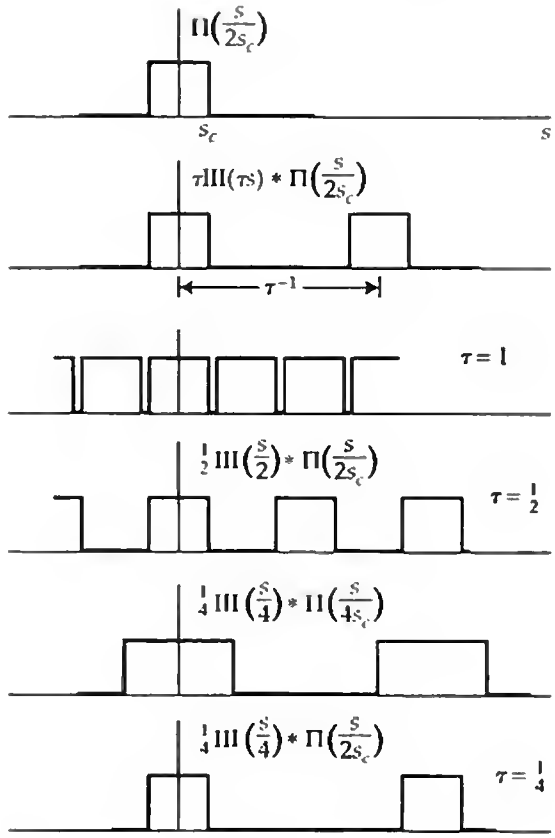


Fig. 10.5 **Numerical procedure for achieving the desired filter characteristic $\Pi(s)$.**

$\frac{1}{4}\text{III}(s/4) * \Pi(s/2)$, which when multiplied by $\frac{1}{2}\text{III}(s/2) * \Pi(s)$ gives the bottom line of Fig. 10.5. In other words, repeated application of the process has pushed down more of the outer islands of response. The same result is obtained by taking $\tau = \frac{1}{4}$ initially.

To summarize, approximate rectangular filtering with a cutoff at s_c is carried out by reading off $f(x)$ at half the critical sampling interval (that is, at intervals of $\frac{1}{4}s_c^{-1}$) and taking the convolution (more precisely, the serial product) with $2s_c$ sinc $2s_cx$, where $2s_cx$ assumes all half-integral values, including 0. This filtering array for use in the function domain (see Table 10.2) contains precisely the values tabulated for interpolation plus interleaved zeros and a central value of unity.

SMOOTHING BY RUNNING MEANS

Convolving a function $f(x)$ with a rectangle function of width W results in the transform $F(s)$ being subjected to a form of low-pass filtering that is far from having the sharp cutoff discussed above. We understand that the transfer function

TABLE 10.2
Array for approximate rectangular filtering

	⋮
	−0.0909
	0
	0.1273
	0
	−0.2122
	0
	0.6366
→	1.0000
	0.6366
	0
	−0.2122
	0
	0.1273
	0
	−0.0909
	⋮

associated with convolution with $W\Pi(x/W)$, given by $F(s) = \operatorname{sinc} Ws$, has a pass band of equivalent width $1/W$, or a 3 dB bandwidth of $0.8859/W$, but that the fall-off in transmission is not at all sharp enough for many purposes. Moreover, the transmission does not descend gracefully to zero but overshoots and oscillates rather strongly. How to make a trade-off between sharpness of cutoff and mildness of the sidelobes, to borrow a convenient term from antenna practice, will now be discussed, starting with the rectangular convolving function as a basis for comparison. However, we look at the convolving operation as a numerical one to be applied to discrete data, in the spirit of Chapter 3 and the exercises there.

Suppose that $W = 12$, a case that can be thought of as generating 12-month running means from 12 data samples one month apart. Instead of $\Pi(x/W)$ we deal with $\frac{1}{12}\sum\delta(x - n)$, for $n = \pm\frac{1}{2}, \pm 1\frac{1}{2}, \ldots \pm 5\frac{1}{2}$. The associated transfer function will be like sinc $12s$ for small s, but will be replicated at unit intervals in the s-domain, as follows from the transform of $\frac{1}{12}\text{III}(x - \frac{1}{2})\Pi(x/12)$, namely $[e^{-i\pi s}\text{III}(s)] * \operatorname{sinc} 12s$. The overlapping sinc functions can be simplified by transforming the impulses $\sum\delta(x - n)$ pair by pair to get $\sum 2\cos 2\pi ns$, $n = \frac{1}{2}, 1\frac{1}{2}, 2\frac{1}{2}, \ldots 5\frac{1}{2}$. To condense further, transform impulse by impulse to get the geometric series

$$\frac{1}{12}(e^{i2\pi 5\frac{1}{2}s} + e^{i2\pi 4\frac{1}{2}s} + \ldots + e^{-i2\pi 4\frac{1}{2}s} + e^{-i2\pi 5\frac{1}{2}s}),$$

where the ratio r of each term to the preceding one is $e^{-i2\pi s}$ and the first term is $a = (\frac{1}{12})\exp(i2\pi 5\frac{1}{2}s)$. The expression $a(1 - r^n)/(1 - r)$ for the sum of a geometric series of n terms then gives

$$\frac{1}{12}e^{i2\pi 5\frac{1}{2}s}(1 - r^{12})/(1 - r) = \frac{1}{12}e^{i2\pi 5\frac{1}{2}s}(1 - e^{-i2\pi 12s})/(1 - e^{-i2\pi s})$$
$$= \frac{1}{12}(e^{i2\pi 6s} - e^{-i2\pi 6s})/(e^{-i2\pi\frac{1}{2}s} - e^{i2\pi\frac{1}{2}s}) = \frac{1}{12}\frac{\sin 2\pi 6s}{\sin 2\pi\frac{1}{2}s}.$$

In general, for N equispaced impulses of strength $1/N$ this expression becomes $(\sin N\pi s)/(N \sin \pi s)$, which is convenient for computing and is shown in Fig. 10.6.

It has been standard practice in meteorology to calculate annual running means using 13 coefficients one month apart instead of 12 but to reduce the first and last to half strength, keeping the sum of the strengths at 12; the transition to zero is smoother. In this case the transfer function becomes $\sin N\pi s/(N \tan \pi s)$.

The upper curve of Fig. 10.6 is recognizable as the antenna field pattern of an array of 12 well-spaced small antennas, while the square of the same curve is the optical diffraction pattern of a row of 12 pinholes.

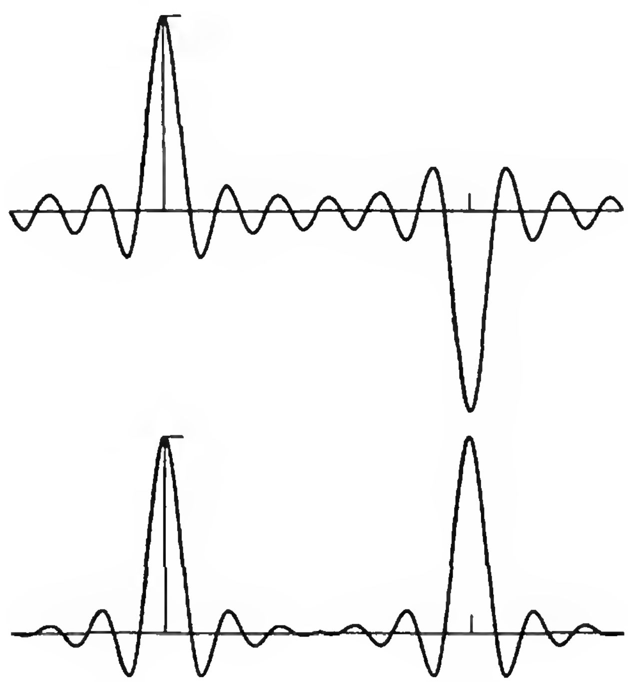


Fig. 10.6 The transfer function for 12 coefficients of strength $\frac{1}{12}$ (top curve). With 13 coefficients, those at the ends being of half strength, the sidelobes are somewhat reduced (bottom) and the first order "grating lobes" are not reversed in phase.

UNDERSAMPLING

Suppose that $f(x)$ (see Fig. 10.7*a*) is read off at intervals corresponding to a desired cutoff for rectangular filtering. Then this set of values (see Fig. 10.7*b*) defines a band-limited function $g(x)$ (see Fig. 10.7*c*) that has a cutoff spectrum of the desired extent and may at first sight appear to be a product of rectangular filtering. But the process is not the same as rectangular filtering, since the result depends on high-frequency components in $f(x)$; for example, one of the sample values may fall at the peak of a narrow spike; furthermore, the *phase* of the coarse sampling points will clearly affect the result. However, the effect may often be a good approximation to rectangular filtering.

By examining the process in terms of Fourier transforms, we see that the band-limited function $g(x)$ derived from too-coarse sampling contains contributions from high-frequency components of $f(x)$, impersonating low frequencies in a way

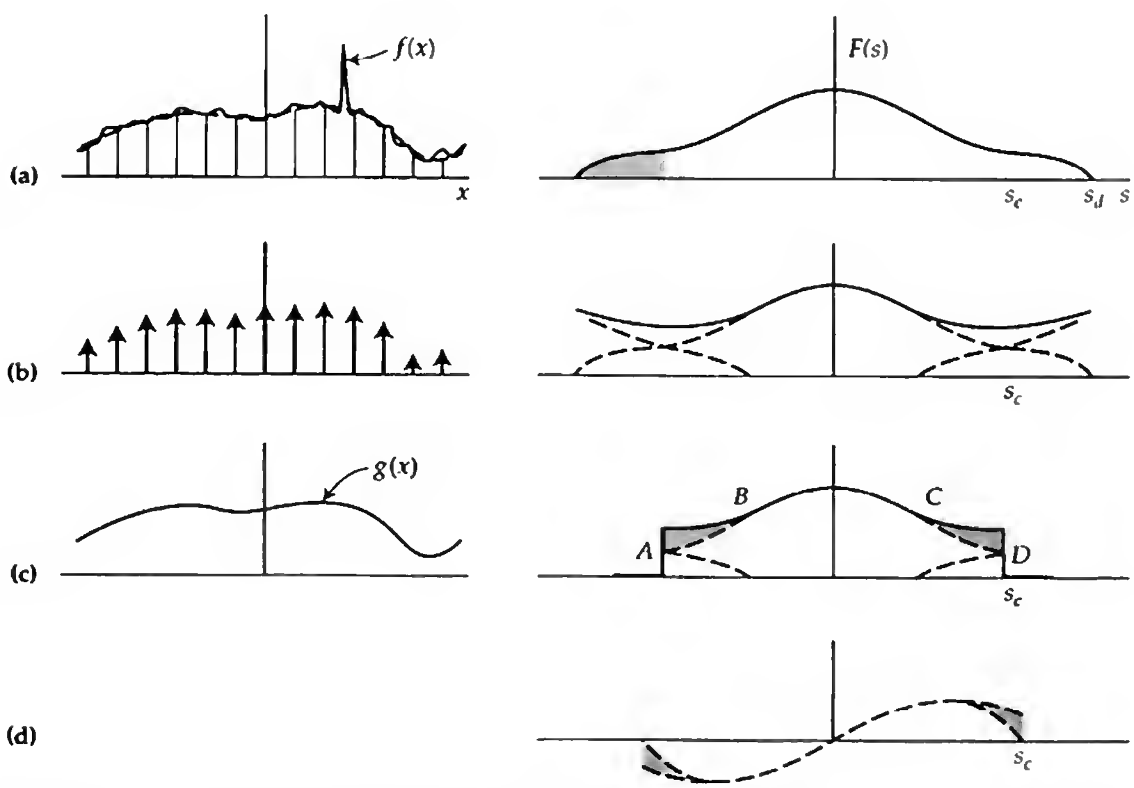


Fig. 10.7 A band-limited function $g(x)$ derived from $f(x)$ by undersampling; high-frequency components of $f(x)$ shift inside the band limits. Shaded areas indicate high frequencies masquerading as low.

that looks like reflection of the high-frequency part in the line $s = s_c$. The effect has been referred to as "aliasing." Closer scrutiny of Fig. 10.7(*b*), taking into account the omitted imaginary part of the transform, will reveal that the spurious low positive frequencies derive from the shaded *negative*-frequency tail of $F(s)$. If this high-frequency tail is not too important, then the coarse sampling procedure gives a fair result.

While the effect of undersampling is to reinforce the even part of the spectrum, Fig. 10.7*d* shows that the odd part is diminished. It follows that $g(x)$ will be evener than $f(x)$, which is, of course, to be expected, for the sampling procedure discriminates against the (necessarily odd) components with zeros at the sampling points.

ORDINATE AND SLOPE SAMPLING

Let $f(x)$ be a band-limited function that is fully specified by ordinates at a spacing of 0.5; but suppose that only IIIf is given; that is, only every second ordinate is given. Then from the overlapping islands (see Fig. 10.8) composing $\overline{\text{III}f}$, it would not be possible to recover $F(s)$ or, consequently, $f(x)$. But if further partial information were given, it might become possible.

If the slope is given, in addition to the ordinate, recovery proves to be possible. Thus, given IIIf and IIIf', one can express $f(x)$ as a combination of linear func-

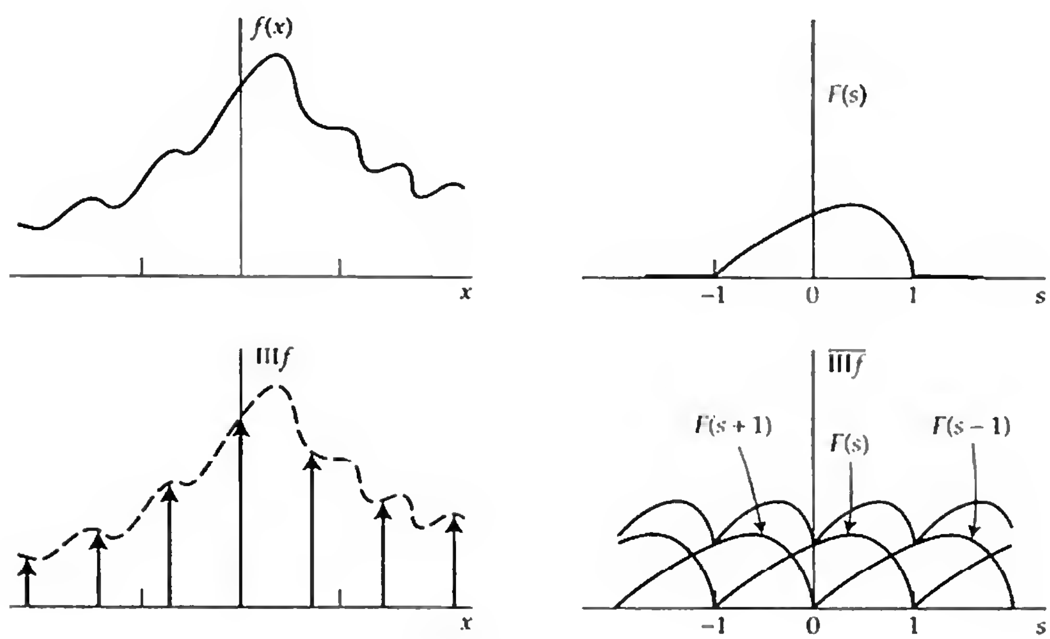


Fig. 10.8 **Ordinate sampling at half the rate necessary for full definition.**

tionals of IIIf and IIIf':

$$f(x) = a(x) * (\mathrm{III}f) + b(x) * (\mathrm{III}f'),$$

where $a(x)$ and $b(x)$ are solving functions that have to be found. Just as the sinc function, which is the solving function for ordinary sampling, must be zero at all its sample points save the origin, where it must be unity, so must $a(x)$. And $b(x)$ must be zero at all sample points but have unit slope at the origin.

To find $a(x)$ and $b(x)$, note that in the interval $-1 \leq s \leq 1$

$$\overline{\mathrm{III}f} = F(s+1) + F(s) + F(s-1)$$

and that $$\overline{\mathrm{III}f'} = i2\pi(s+1)F(s+1) + i2\pi sF(s) + i2\pi(s-1)F(s-1).$$

These two equations can be solved for $F(s)$, for although there appear to be three unknowns, namely, $F(s+1)$, $F(s)$, $F(s-1)$, in fact, for any value of s, one of them is always known to be zero. Thus for positive s we have $F(s+1) = 0$, and on eliminating $F(s-1)$ we have

$$i2\pi F(s) = \overline{\mathrm{III}f'} - i2\pi(s-1)\overline{\mathrm{III}f},$$

and for negative s we have

$$-i2\pi F(s) = \overline{\mathrm{III}f'} - i2\pi(s+1)\overline{\mathrm{III}f}.$$

For all s, both positive and negative, we have concisely

$$F(s) = \frac{i}{2\pi}\Lambda'(s)\overline{\mathrm{III}f'} + \Lambda(s)\overline{\mathrm{III}f}.$$

Hence $$f(x) = \operatorname{sinc}^2 x * (\mathrm{III}f) + x\operatorname{sinc}^2 x * (\mathrm{III}f').$$

The solving functions are

$$a(x) = \operatorname{sinc}^2 x$$
$$b(x) = x\operatorname{sinc}^2 x,$$

and the convolution integrals reduce to a sum of spaced a's and b's of suitable amplitudes,

$$\begin{aligned} f(x) &= \sum_{-\infty}^{\infty} f(n)a(x-n) + \sum_{-\infty}^{\infty} f'(n)b(x-n) \\ &= \sum_{-\infty}^{\infty} f(n)\operatorname{sinc}^2(x-n) + \sum_{-\infty}^{\infty} f'(n)(x-n)\operatorname{sinc}^2(x-n). \end{aligned}$$

We see that each $a(x-n)$ is zero at all sampling points (integral values of x) save where $x = n$, and that each has zero slope at all sampling points (see Fig. 10.9a). Each $b(x-n)$ has zero ordinate at all sampling points and likewise zero slope, save where $x = n$ (see Fig. 10.9b).

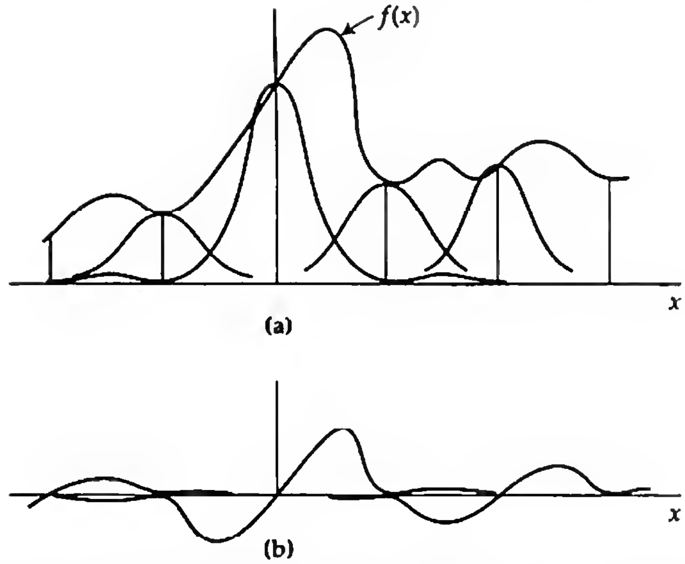


Fig. 10.9 (a) The ordinate-dependent constituents $f(n)a(x - n)$; (b) The slope-dependent constituents $f'(n)b(x - n)$.

INTERLACED SAMPLING

As in the previous example, let IIIf represent every second ordinate of the set that is necessary to specify $f(x)$, and let a supplementary set $\mathrm{III}(x - a)f(x)$, interlaced with the first as illustrated in Fig. 10.10, also be available. Will it be possible to reconstitute $f(x)$? It is known that equispaced samples, separated by just more than the critical interval, do not suffice, and in the case of interlaced sampling every second jump exceeds this critical interval. On the other hand, it has been shown that ordinate-and-slope sampling suffices, and this is clearly equivalent to extreme interlacing as a approaches zero.

If there is a solution, it should be in the form of a sum of two linear functionals of IIIf and $\mathrm{III}_a f$, where $\mathrm{III}_a \equiv \mathrm{III}(x - a)$:

$$f(x) = a(x) * (\mathrm{III}f) + b(x) * (\mathrm{III}_a f).$$

The solving function $a(x)$ must be equal to unity at $x = 0$, and zero at all other sampling points, and $b(x)$ must be the mirror image of $a(x)$; that is, $b(x) = a(-x)$.

In the interval $-1 < s < 1$,

$$\overline{\mathrm{III}f} = F(s + 1) + F(s) + F(s - 1)$$

and

$$\overline{\mathrm{III}_a f} = e^{i2\pi a}F(s + 1) + F(s) + e^{-i2\pi a}F(s - 1).$$

For positive s we have $F(s + 1) = 0$, and on eliminating $F(s - 1)$ we have

$$F(s) = -\frac{e^{-i2\pi a}}{1 - e^{-i2\pi a}}\overline{\mathrm{III}f} + \frac{1}{1 - e^{-i2\pi a}}\overline{\mathrm{III}_a f},$$

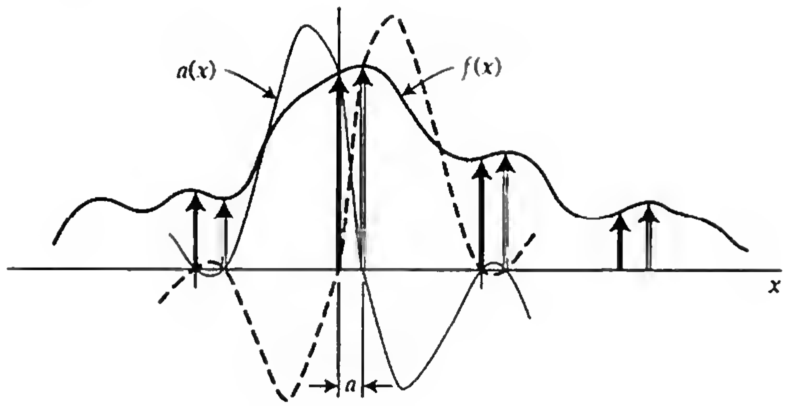


Fig. 10.10 **Interlaced samples.**

and for negative s we have

$$F(s) = -\frac{e^{i2\pi a}}{1 - e^{i2\pi a}}\,\overline{\mathrm{III}f} + \frac{1}{1 - e^{i2\pi a}}\,\overline{\mathrm{III}_a f}.$$

For all s we have

$$F(s) = A(s)\overline{\mathrm{III}f} + A^*(s)\overline{\mathrm{III}_a f},$$

where

$$A(s) = \begin{cases} -\dfrac{e^{i2\pi a}}{1 - e^{-i2\pi a}} & 0 < s < 1 \\ -\dfrac{e^{i2\pi a}}{1 - e^{i2\pi a}} & -1 < s < 0 \\ 0 & |s| > 1 \end{cases}$$

$$= \tfrac{1}{2}\,\Pi\left(\frac{s}{2}\right) + \tfrac{1}{2} i \cot a\pi\,\Lambda'(s).$$

Hence

$$a(x) = \operatorname{sinc} 2x - (\pi \cot a\pi) x \operatorname{sinc}^2 x,$$

as graphed in Fig. 10.10.[1]

It may seem strange that the equidistant samples may be regrouped in pairs, even to the extreme of close spacing. However, it is also possible to bunch the samples in groups of any size separated by such wide intervals as maintain the original average spacing. The bunching may be indefinitely close; see Linden for a proof that the ordinates and first n derivatives, at points spaced by $n + 1$ times

---

[1] For a discussion of sampling theorems see Linden, (1959). In this paper the solving function $a(x)$ is given in the form

$$a(x) = \frac{\cos(2\pi x - a\pi) - \cos a\pi}{2\pi x \sin a\pi}.$$

The numerator is a cosine wave so displaced that it has zeros at $x = n$ and $x = n + a$, but the zero at $x = 0$ is counteracted by the vanishing of the denominator in such a way that $a(0) = 1$.

the usual spacing, suffice to specify a band-limited function. In the limit, as n approaches infinity, the formula for reconstituting the function becomes the Maclaurin series. This introduces doubt of the practical applicability of higher-order sampling theorems, for it is well known that the Maclaurin series

$$f(0) + xf'(0) + \frac{x^2}{2!}f''(0) + \ldots + \frac{x^n}{n!}f^{(n)}(0)$$

does not usually converge to $f(x)$. (Consider the functions $\Pi(ax)$, which, for different values of a, all have the same Maclaurin series.)

In practice, higher-order sampling breaks down at some point because small amounts of noise drastically affect the determination of high-order derivatives or finite differences. In the total absence of noise, trouble would still be expected to set in at some stage because of the impossibility of ensuring perfectly band-limited behavior. In applications of sampling theorems, the claim of a given function to be band-limited must always be scrutinized, and any error resulting must be estimated.

SAMPLING IN THE PRESENCE OF NOISE

Suppose that the samples $f(n)$ of a certain function $f(x)$ cannot be obtained without some error being made; that is, the observable quantity is

$$f(n) + \text{error}.$$

Then when an attempt is made to reconstruct $f(x)$ from the observed samples by applying the same procedure used for true samples, the reconstructed values will differ from the true values of $f(x)$.

Consider the case of midpoint interpolation, supposing that samples have been taken at $x = \pm\frac{1}{2}, \pm 1\frac{1}{2}, \pm 2\frac{1}{2}, \ldots$. Then

$$f(0) = 0.6366[f(\tfrac{1}{2}) + f(-\tfrac{1}{2})] - 0.2122[f(1\tfrac{1}{2}) + f(-1\tfrac{1}{2})] + 0.1273[f(2\tfrac{1}{2}) + f(-2\tfrac{1}{2})] - \ldots.$$

The errors affecting $f(\frac{1}{2})$ and $f(-\frac{1}{2})$ will have the greatest effect on $f(0)$; each error is reduced by 0.6366, and the resulting net error at $x = 0$, due to the errors at the nearest two sampling points, could be anywhere from zero, if the two errors happened to cancel, up to 1.27 times either error. Clearly, the error involved in using the sampling theorem to interpolate will be of the same order of magnitude as the errors affecting the data. More than this could not be expected, and so it may be concluded that the interpolating procedure is tolerant to the presence of noise.

It is not the purpose here to go into statistical matters, but a simple result should be pointed out that arises when the errors are of such a nature that they are independent from one sample to the next, and all come from a population with zero mean value and a certain variance σ^2. Then the variance of the error contributed at $x = 0$ by the error at $x = \frac{1}{2}$ is $(0.6366)^2\sigma^2$, and the variance of the

total error at $x = 0$ due to the errors at all the sampling points is

$$[\ldots + (0.1273)^2 + (0.2122)^2 + (0.6366)^2 + (0.6366)^2 + (0.2122)^2 + (0.1273)^2 + \ldots]\sigma^2.$$

Now the terms of the series within the brackets are values of $\operatorname{sinc}^2 x$ at unit intervals of x and hence add up to unity. Therefore, in this simple error situation, the interpolated value is subject to precisely the same error as the data.

Now we apply the same reasoning to interlaced sampling, especially to the extreme situation where the narrow interval is small compared with the wide one, that is, where $a \ll 1$. At the point $x = \frac{1}{2}a + \frac{1}{2}$, which is in the middle of the wide interval, the four nearest sample values enter with coefficients

$$a(\tfrac{1}{2}a + \tfrac{1}{2}),\ b(-\tfrac{1}{2}a + \tfrac{1}{2}),\ a(\tfrac{1}{2}a - \tfrac{1}{2}),\ b(-\tfrac{1}{2}a - \tfrac{1}{2}).$$

The first and last of these are positive and the others negative, and the presence of the factor $\cot a\pi$ in the formula for $a(x)$ shows that the numerical values may be large. Thus the interpolated value may result from the cancellation of large terms, and the total error may be large.

In another way of looking at this, a pair of terms

$$f(0)a(x) + f(a)b(x - a)$$

may be reexpressed in the form

$$[a(x) + b(x - a)]\left[\frac{f(0) + f(a)}{2}\right] + \frac{b(x - a) - a(x)}{2}[f(a) - f(0)].$$

Here the first term represents the mean of a sample pair multiplied by a certain coefficient, and the second represents the difference between two close-spaced samples multiplied by a certain other coefficient. In the limit as $a \to 0$, the solving functions for the ordinate-and-slope sampling theorem would result. Now the coefficient of the difference term can be large; for example, when $a < 0.2$, the value adopted in the illustration of interlaced sampling, the coefficient exceeds unity. Thus errors in the difference terms may be amplified.

It thus appears that interlaced sampling, where the sample spacing is alternately narrow and wide, is not tolerant to errors, and therefore the magnitude of the errors would have to be estimated carefully in an application. In a full study it would be essential to take account of any correlation between the errors in successive samples since it is clear that the error in $f(a) - f(0)$ would be reduced if both $f(a)$ and $f(0)$ were subject to about the same error.

FOURIER SERIES

It is well known that a periodic waveform, such as the acoustical waveform associated with a sustained note of a musical instrument, is composed of a fundamental and harmonics. Exploration of such an acoustical field by means of tun-

able resonators reveals that the energy is concentrated at frequencies which are integral multiples of the fundamental frequency. There is nothing here that should exclude this case from treatment by the Fourier transform methods so far used. However, insistence on strict periodicity, a physically impossible thing, will clearly lead to an impulsive spectrum and to the refined considerations that are needed in connection with impulses. We now proceed to do this, using the *shah* symbol for convenient handling of the sets of impulses that arise, in connection both with the replication inherent in periodicity and with the sampling associated with harmonic spectra.

The Fourier series will be exhibited as an extreme situation of the Fourier transform, even though the opposite procedure, taking the Fourier series as a point of departure for developing the Fourier transform, is traditional.

For reference let it be stated that the Fourier series associated with the periodic function $g(x)$, with frequency f and period T, is

$$a_0 + \sum_1^\infty (a_n \cos 2\pi n f x + b_n \sin 2\pi n f x),$$

where
$$a_0 = \frac{1}{T}\int_{-\frac{1}{2}T}^{\frac{1}{2}T} g(x)\,dx$$

$$a_n = \frac{2}{T}\int_{-\frac{1}{2}T}^{\frac{1}{2}T} g(x)\cos 2\pi n f x\,dx$$

$$b_n = \frac{2}{T}\int_{-\frac{1}{2}T}^{\frac{1}{2}T} g(x)\sin 2\pi n f x\,dx.$$

It is necessary for $g(x)$ to have been chosen so that the integrals exist; otherwise $g(x)$ is arbitrary.

The purpose of a good deal of theory dealing with the Fourier series has been to show that the series associated with a periodic function $g(x)$ does in fact often converge, and furthermore, that when it converges, it often converges to

$$\tfrac{1}{2}[g(x+0) + g(x-0)].$$

The rigorous development of this topic was initiated by Dirichlet in 1829, following a controversial period dating back to D. Bernoulli's success in 1753 in expressing the form of a vibrating string as a series

$$y = A_1 \sin x \cos at + A_2 \sin 2x \cos 2at + \dots.$$

Euler, who had been working on this problem and had just obtained the general solution in terms of traveling waves, said that if Bernoulli was right, an arbitrary function could be expanded as a sine series. This, he said, was impossible. In 1807, when Fourier made this same claim in his paper presented to the Paris Academy, Lagrange rose and said it was impossible. This exciting subject led to many important developments in pure mathematics, including the invention of the Riemann integral. It must be remembered that the expressions for the Fourier constants a_0, a_n, b_n were given long before modern analysis developed.

For the present purpose let us take $T = 1$ and note that the complex constant $a_n - ib_n$ is related to the one-period segment $g(x)\Pi(x)$ by the Fourier transform

$$\begin{aligned} a_n - ib_n &= 2\int_{-\frac{1}{2}}^{\frac{1}{2}} g(x)e^{-i2\pi nx}\,dx \\ &= 2\int_{-\infty}^{\infty} g(x)\Pi(x)e^{-i2\pi nx}\,dx = 2\,F(n). \end{aligned}$$

We now recover this result by considering the Fourier transform of a periodic function.

Let $f(x)$ be a function that possesses a regular Fourier transform $F(s)$ (see Fig. 10.11). Then its convolution with the replicating symbol $\mathrm{III}(x)$ will be the periodic function $p(x)$, defined by

$$p(x) = \mathrm{III}(x) * f(x) = \sum_{-\infty}^{\infty} f(x - n) \qquad n \text{ integral.}$$

The convergence of the summation is guaranteed by the absolute integrability of $f(x)$, which was requisite for the possession of a Fourier transform. The period of $p(x)$ is unity; that is,

$$p(x + 1) = p(x)$$

for all x.

There will be no regular Fourier transform for $p(x)$ because

$$\int_{-\infty}^{\infty} |p(x)|\,dx$$

cannot converge, save in the trivial case where $p(x)$ is identically zero. Hence we multiply $p(x)$ by a factor $\gamma(x)$ that dies out to zero for large values of x, both pos-

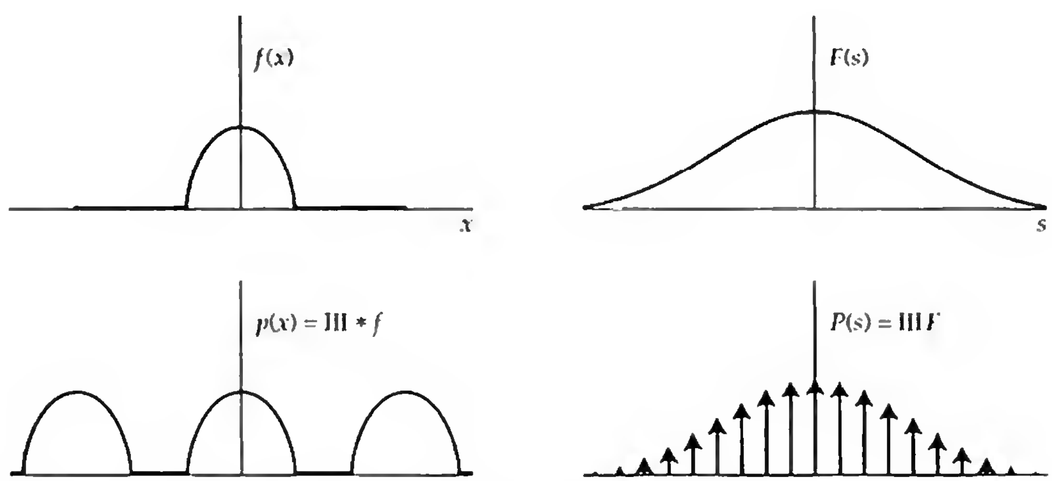


Fig. 10.11 **The transform in the limit of a periodic function.**

itive and negative. In effect we are bringing the strictly periodic function $p(x)$ back to the realm of the physically possible, but only barely so (see Fig. 10.12). Let

$$\gamma(x) = e^{-\pi\tau^2x^2};$$

then the Fourier transform $\Gamma(s)$ of $\gamma(x)$ will be given by

$$\Gamma(s) = \tau^{-1}e^{-\pi s^2/\tau^2}.$$

The function $\gamma(x)p(x)$ will possess a Fourier transform,

$$\begin{aligned}\gamma(x)p(x) &= \gamma(x)\sum_{-\infty}^{\infty} f(x-n)\\ &\supset \Gamma(s) * \sum_{-\infty}^{\infty} e^{-i2\pi ns}F(s)\\ &= \sum_{-\infty}^{\infty} F(n)\Gamma(s-n).\end{aligned}$$

In *shah*-symbol notation,

$$\sum_{-\infty}^{\infty} F(n)\Gamma(s-n) = \Gamma(s) * [\text{III}(s)F(s)].$$

Now let

$$P(s) = \text{III}(s)F(s).$$

This entity $P(s)$ is a whole set of equidistant impulses of various strengths, as given by samples of $F(s)$ at integral values of s, and has the property that

$$\gamma p \supset \Gamma * P.$$

By the convolution theorem we see that P is a suitable symbol to assign as the Fourier transform of $p(x)$, and indeed as $\tau \to 0$ and γp runs through a sequence of functions having p as limit, the sequence $\Gamma * P$ defines an entity P of such a type that, as agreed, we may call $p(x)$ and $P(s)$ a Fourier transform pair in the limit.

Taking $f(x)$ to be the one-period segment $g(x)\Pi(x)$, we see that the spectrum of a periodic function is a set of impulses whose strengths are given by equidis-

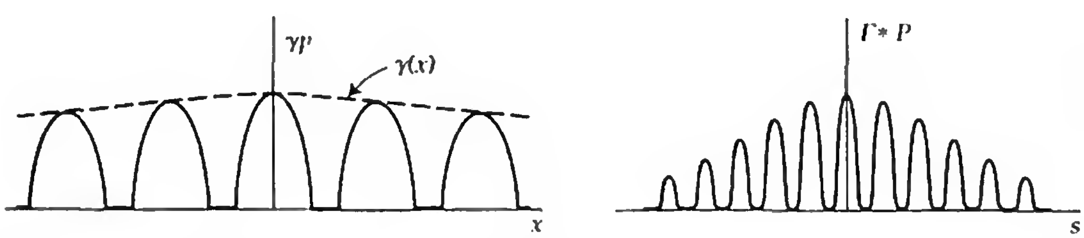


Fig. 10.12 **The convergence factor γ applied to a periodic function $p(x)$ renders its infinite integral convergent while the convolvent Γ removes infinite discontinuities from a line spectrum $P(s)$.**

tant samples of $F(s)$, the Fourier transform of the one-period segment. Now

$$\begin{aligned}
\int_{-\infty}^{\infty} \mathrm{III}(s)F(s)e^{+i2\pi sx}\,ds &= \sum_{-\infty}^{\infty} \int_{n-0}^{n+0} F(s)e^{+i2\pi sx}\,ds \\
&= \sum_{-\infty}^{\infty} F(n)e^{+i2\pi nx} \\
&= F(0) + \sum_{1}^{\infty} [F(n)e^{+i2\pi nx} + \text{conjugate}] \\
&= F(0) + 2\sum_{1}^{\infty} (\mathrm{Re}\,F \cos 2\pi nx - \mathrm{Im}\,F \sin 2\pi nx) \\
&= a_0 + \sum_{1}^{\infty} (a_n \cos 2\pi nx + b_n \sin 2\pi nx)
\end{aligned}$$

if $a_n - ib_n = 2F(n)$. And this is precisely the value that was quoted earlier for the complex coefficient $a_n - ib_n$.

The fact that rigorous deliberations on Fourier series generally are more complex than those encountered with the Fourier integral is essentially connected with the infinite energy of periodic functions (the integrals of which are not absolutely convergent). It is therefore very natural physically to regard a periodic *function* as something to be approached through functions having finite energy, and to consider a *line spectrum* as something to be approached via continuous spectra with finite energy density. The strange thing is that the physically possible functions and spectra are often presented as elaborations of the physically impossible. Some people cannot see how a *line* spectrum, no matter how closely the lines are packed, can ultimately become a continuous spectrum. This order of presentation is inherited from the historical precedence of the theory of trigonometric series, and runs as follows (in our terminology). The periodic function $\mathrm{III} * f$ has a line spectrum $\mathrm{III}F$ (set of Fourier coefficients). If the repetition period is lengthened to τ, the lines of the spectrum are packed τ times more closely and are τ times weaker (note compensating change in ordinate scale in Fig. 10.13). Now let the period become infinite, so that the pulse f does not recur. Then the trigonometric sum which represented its periodic predecessors passes into an infinite integral, and the finite integral which specified the series coefficients does likewise. These two integrals are the plus-i and minus-i Fourier integrals.

On the view described here line spectra and periodic functions are regarded as included in the theory of Fourier transforms, to be handled exactly as other transforms by means of the III symbology, and with the same caution accorded to other transforms in the limit.

As assumed in Fig. 10.13 an infinite array of input impulses at spacing τ transforms into impulses at spacing τ^{-1} but not of unit strength. If $\tau > 1$, as illustrated, then the spacing of the impulses of the transform is less than unity and their strength is greater than unity. Thus

$$\frac{1}{|\tau|}\,\mathrm{III}\left(\frac{x}{\tau}\right) \supset \mathrm{III}(\tau s).$$

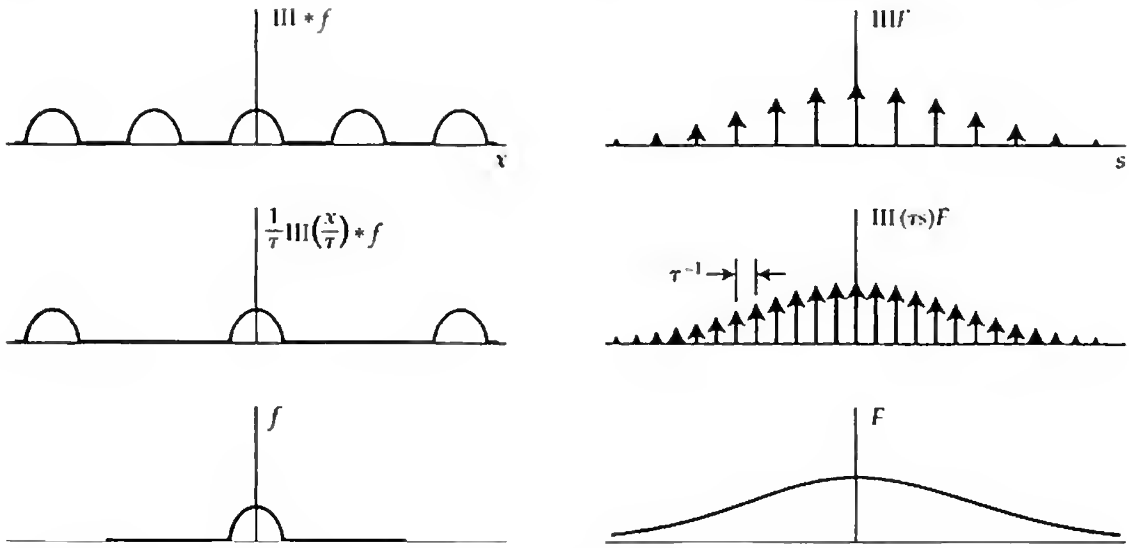


Fig. 10.13 Deriving the Fourier integral from the Fourier series.

If we write $\text{III}(x/\tau) \supset |\tau|\text{III}(\tau s)$ then the right-hand side consists of unit impulses (spaced τ^{-1}) while $\text{III}(x/\tau)$ has stronger impulses (taking $\tau > 1$). These relations result from the similarity theorem. One can understand that, when a train of brief pulses is stretched, each pulse is widened and gains area; the corresponding impulses gain proportionately in strength.

Gibbs phenomenon. One of the classical topics of the theory of Fourier series may be studied profitably from the point of view that we have developed. In situations where periodic phenomena are analyzed to determine the coefficients of a Fourier series, which is then used to predict, it is a matter of practical importance to know how many terms of the series to retain. Various considerations enter into this, but one of them is the phenomenon of overshoot associated with discontinuities, or sharp changes, in the periodic function to be represented.

It is quite clear that by omitting terms beyond a certain limiting frequency we are subjecting the periodic function $p(x)$ to low-pass filtering. Thus, if the fundamental frequency is s_0, and frequencies up to ns_0 are retained, it is as though the spectrum had been multiplied by a rectangle function $\Pi(s/2s_c)$, where s_c is a cutoff frequency between ns_0 and $(n+1)s_0$. It makes no difference precisely where s_c is taken; for convenience it may be taken at $(n+\frac{1}{2})s_0$. Multiplication of the spectrum by $\Pi[s/(2n+1)s_0]$ corresponds to convolution of the original periodic function $p(x)$ with $(2n+1)s_0 \operatorname{sinc}[(2n+1)s_0x]$. Therefore, when the series is summed for terms up to frequencies ns_0 only, the sum will be

$$p(x) * (2n+1)s_0 \operatorname{sinc}[(2n+1)s_0x].$$

The convolving function has unit area, so in places where $p(x)$ is slowly varying, the result will be in close agreement with $p(x)$.

We now wish to study what happens at a discontinuity, and so we choose a periodic function which is equal to sgn x for a good distance to each side of $x = 0$ (see Fig. 10.14). What it does outside this range will not matter, as long as it is periodic, because we are going to focus attention on what happens near $x = 0$. Near $x = 0$ the result will be approximately

$$(2n + 1)s_0 \operatorname{sinc}[(2n + 1)s_0 x] * \operatorname{sgn} x.$$

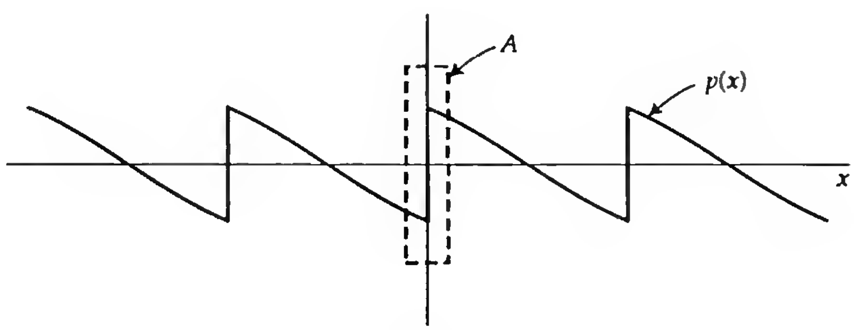
We know that

$$\operatorname{sinc} x * \operatorname{sgn} x = 2\int_0^x \operatorname{sinc} t \, dt,$$

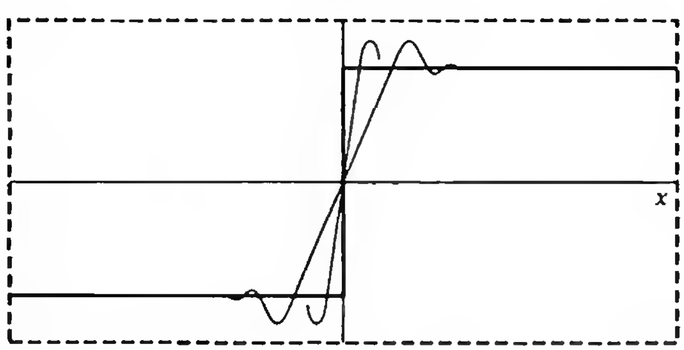
a function closely related to the sine integral Si(x). In fact

$$2\int_0^x \operatorname{sinc} t \, dt = \frac{2}{\pi}\operatorname{Si}(\pi x).$$

This function oscillates about -1 for large negative values of x, oscillates with increasing amplitude as the origin is approached, passes through zero at $x = 0$, shoots up to a maximum value of 1.18, and then settles down to decaying oscillations about a value of $+1$. If we change the scale factors of the sinc function, compressing it by a factor $N = (2n + 1)s_0$, and strengthening it by a factor N, so



(a)



(b)

Fig. 10.14 **(a) A periodic function $p(x)$; (b) an enlargement of area A.**

as to retain its unit area, then convolution with sgn x will result in oscillations about -1 and then about $+1$ that are faster but have the same amplitude, that is,

$$N \operatorname{sinc} Nx * \operatorname{sgn} x = \frac{2}{\pi} \operatorname{Si}(N\pi x).$$

The overshoot, amounting to 9 percent of the amount of the discontinuity, remains at 9 percent, but the maximum is reached nearer to the discontinuity. The same applies to the minimum that occurs on the negative side.

Now we see precisely what happens when a Fourier series is truncated. There is overshoot on both sides of any discontinuity, amounting to about 9 percent, regardless of the inclusion of more and more terms. At any given point to one side of the discontinuity the oscillations decrease indefinitely in amplitude as N increases, so that in the limit the sum of the series approaches the value of the function of which it is the Fourier series (and at the point of discontinuity, the sum of the series approaches the midpoint of the jump). In spite of this, the maximum departure of the sum of the series from the function, i.e., the error, remains different from zero, and as the maximum moves in close to the step, it approaches the precise value of 9 percent that was derived for sgn x, because the parts of the function away from the discontinuity have indeed become irrelevant.

This behavior is reminiscent of that of $x\,\delta(x)$, which is zero for all x, even though sequences defining it have nonvanishing maxima and minima.

Finite Fourier transforms. In problems where the range of the independent variable is not from $-\infty$ to ∞, advantages accrue from the introduction of finite transforms; for example,

$$F(s,a,b) = \int_a^b f(x)e^{-i2\pi xs}\,dx.$$

With such a definition one can work out an inversion formula, a convolution theorem, theorems for the finite transform of derivatives of functions, and so on. As a particular example of an inversion formula we have

$$F_s(s,0,\tfrac{1}{2}) = \int_0^{\frac{1}{2}} f(x) \sin 2\pi xs\,dx,$$

$$f(x) = 4\sum_{s=1}^{\infty} F_s(s,0,\tfrac{1}{2}) \sin 2\pi xs.$$

The right-hand side of the last equation will be recognized as the Fourier series for a periodic function of which the segment in the interval $(0, \frac{1}{2})$ is identical with $f(x)$.

It will be evident that the theory of finite transforms will be the same as the theory of Fourier series and that the principal advantage of their use will lie in the approach. We have seen the convenience of embracing Fourier series within the scope of Fourier transforms through the concept of transforms in the limit, and we shall therefore also include finite transforms. Thus we may write the fore-

going example in terms of ordinary sine transforms as follows:

$$\int_0^{\frac{1}{2}} f(x) \sin 2\pi xs\, dx = 2\int_0^{\infty} [\tfrac{1}{2}\,\Pi(x)f(x)] \sin 2\pi xs\, dx$$

and in the general case

$$\int_a^b f(x)e^{-i2\pi xs}\, dx = \int_{-\infty}^{\infty} \left[\Pi\left(\frac{x - \frac{1}{2}(b + a)}{b - a}\right)f(x)\right]e^{-i2\pi xs}\, dx.$$

In other words, we substitute for integration over a finite range infinite integration of a function which is zero outside the old integration limits.

All the special properties of finite transforms then drop out. For example, the derivative of a function will (in general) be impulsive at the points a and b where it cuts off, and therefore the Fourier transform of the derivative will contain two special terms proportional to the jumps at a and b. It is not necessary to make explicit mention of this property of the transformation when stating the theorem that the Fourier transform of the derivative of a function is $i2\pi s$ times the transform of the function; for example, the Fourier transform of $\Pi'(x)$ is $i2\pi s.\text{sinc}\, s = 2i \sin \pi s$. However, the derivative theorem for finite transforms contains these additive terms explicitly. Thus

$$\int_a^b f'(x)e^{-i2\pi xs}\, dx = i2\pi sF(s,a,b) + f(a)e^{-i2\pi as} - f(b)e^{-i2\pi bs}.$$

Fourier coefficients. If we consider the usual formula for the series coefficients a_n and b_n for a periodic function $p(x)$ of unit period, namely,

$$a_n - ib_n = 2\int_{-\frac{1}{2}}^{\frac{1}{2}} p(x)e^{-i2\pi nx}\, dx,$$

we note that the integral has the form of a finite Fourier transform. Thus in spite of the fact that $p(x)$ is a function of a continuous variable x, whereas the Fourier series coefficients depend on a variable n which can assume only integral values, Fourier transforms as we have been studying them enter directly into the determination of series coefficients. The finite transform can, we know, be expressed as a standard transform of a slightly different function $\Pi(x)p(x)$ as follows:

$$\int_{-\infty}^{\infty} \Pi(x)p(x)e^{-i2\pi nx}\, dx.$$

Our ability to handle transforms can thus be freely applied to the determination of Fourier series coefficients.

As an example consider a periodic train of narrow triangular pulses shown in Fig. 10.15a,

$$\sum_{n=-\infty}^{\infty} \Lambda[10(x - n)].$$

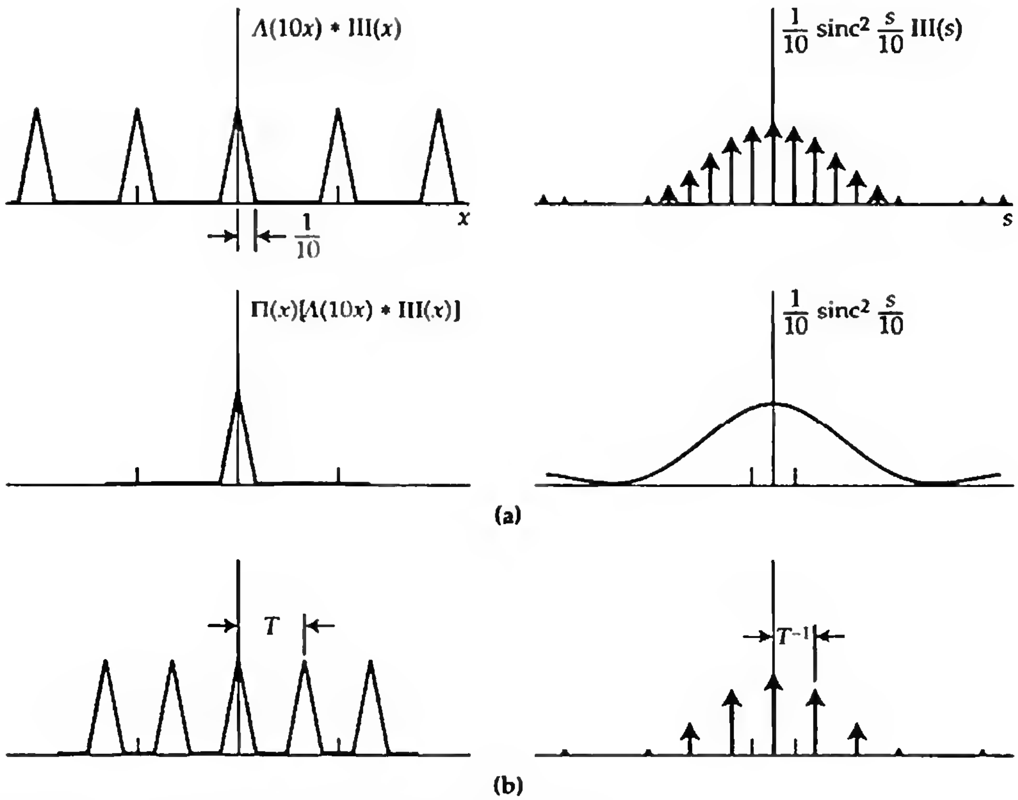


Fig. 10.15 **Obtaining Fourier series coefficients.**

With the aid of the *shah* notation this train can also be expressed in the form

$$\Lambda(10x) * \mathrm{III}(x),$$

whose transform is

$$\frac{1}{10}\,\mathrm{sinc}^2\,\frac{s}{10}\,\mathrm{III}(s).$$

We note that the $\frac{1}{10}\,\mathrm{sinc}^2\,(s/10)$ part of this expression came from

$$\int_{-\infty}^{\infty} \Lambda(10x)e^{-i2\pi sx}\,dx,$$

which, in terms of the periodic train $\Lambda(10x) * \mathrm{III}(x)$, is the same as

$$\int_{-\infty}^{\infty} \Pi(x)[\,\Lambda(10x) * \mathrm{III}(x)]e^{-i2\pi sx}\,dx.$$

Thus by taking the Fourier transform of a single pulse of the train we obtain precisely the expression which arose in connection with the series. It is true that s is a continuous variable while n is not, but

$$\frac{1}{10}\operatorname{sinc}^2\frac{s}{10}\,\mathrm{III}(s)$$

is zero everywhere save at discrete values of s, and at these values the strength of the impulse is equal to the corresponding series coefficient.

In the general case of a function

$$f(x) * \frac{1}{T}\,\mathrm{III}\left(\frac{x}{T}\right)$$

of any period T, the transform

$$F(s)\mathrm{III}(Ts)$$

shows that the coefficients are obtained by reading off the same $F(s)$ at different intervals $s = T^{-1}$. This is illustrated in Fig. 10.15*b*.

IMPULSE TRAINS THAT ARE PERIODIC

A periodic function of x has a transform that is a train of evenly spaced impulses, and vice versa. What if the periodic function is itself composed of evenly spaced impulses, or if an impulse train is itself periodic?

Starting from an original function $f(x)$, let a periodic function $p(x)$ of unit period be generated by convolution with $\mathrm{III}(x)$. Since $p(x)$ has unit period, its transform $P(s)$ will consist of impulses at unit spacing. Now sample $p(x)$ by multiplication with a train of unit impulses with a spacing X that is less than unity to produce a periodic impulse train (Fig. 10.16, bottom left)

$$\hat{f}(x) = [\mathrm{III}(x) * f(x)]\,\frac{1}{X}\,\mathrm{III}\left(\frac{x}{X}\right).$$

By inspection,

$$\hat{f}(x) \supset \hat{F}(s) = [\mathrm{III}(s)F(s)] * \mathrm{III}(Xs).$$

Thus, a regularly sampled periodic function of x (bottom left) produces a structure similar to regular replication of a sampled function (bottom right). It follows that $\hat{f}(x)$ can be expressed in two ways, namely

$$\hat{f}(x) = [\mathrm{III}(x) * f(x)]\,\frac{1}{X}\,\mathrm{III}\left(\frac{x}{X}\right) = \left[f(x)\,\frac{1}{X}\,\mathrm{III}\left(\frac{x}{X}\right)\right] * \mathrm{III}(x).$$

Only the first of these ways is illustrated; there therefore exists another diagram whose top and bottom rows are the same as in Fig. 10.16 but whose central row is replaced by $X^{-1}\mathrm{III}(x/X)f(x)$ on the left and by $\mathrm{III}(Xs) * F(s)$ on the right.

When the discrete Fourier transform is implemented, both the data set and the discrete transform are equivalent to regular impulse trains. Their relationship

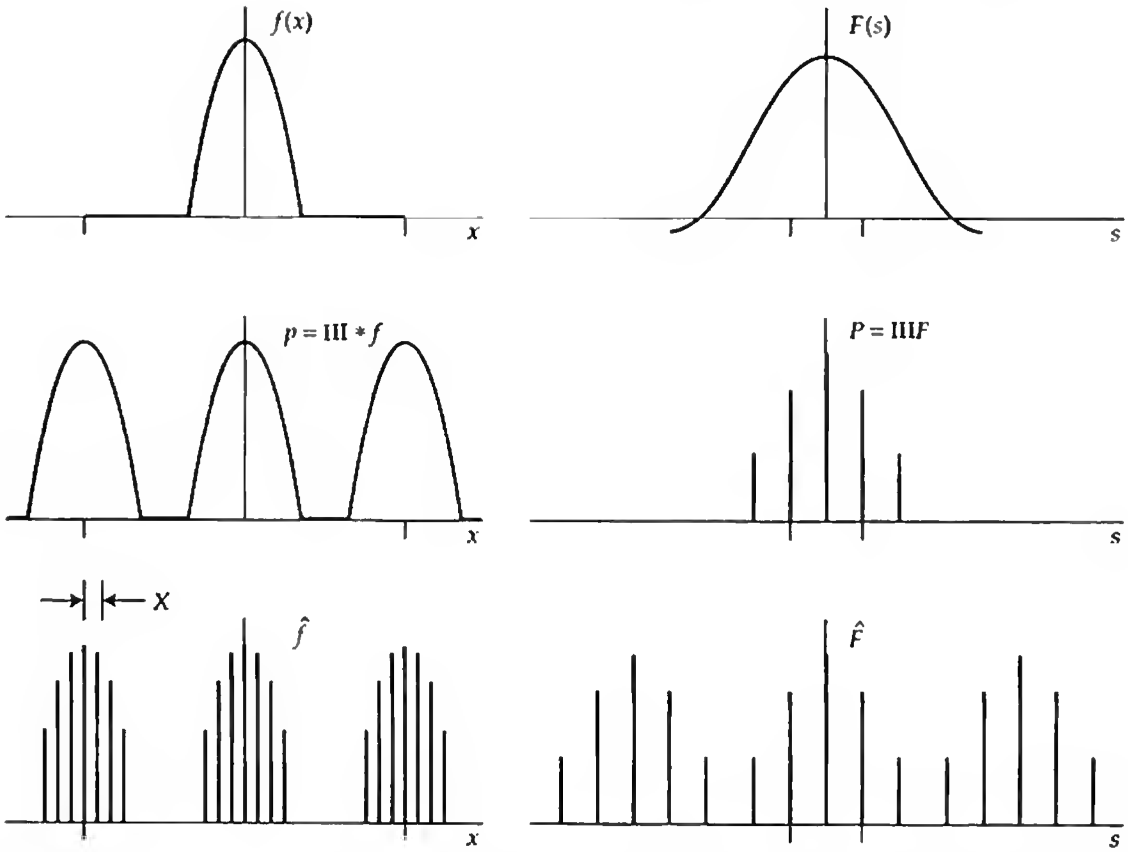


Fig. 10.16 A periodic impulse train (bottom left) generated from $f(x)$ by first replicating (middle left) at unit interval and then sampling at interval X. The corresponding transforms on the right show sampling at unit interval (middle right) followed by replication at interval $1/X$. Ticks on all axes are at ± 1.

is the same as exists between the central clusters of the bottom row. Fig. 10.16 will prove helpful in understanding the phenomena of leakage and aliasing that may cause the coefficients of $\hat{F}(s)$ to differ from the values of $F(s)$, as described in Chapter 11 in the section entitled "Is the Discrete Fourier Transform Correct?"

THE *SHAH* SYMBOL IS ITS OWN FOURIER TRANSFORM

The *shah* symbol $\text{III}(x)$ is defined by

$$\text{III}(x) = \sum_{-\infty}^{\infty} \delta(x - n)$$

and therefore, in accordance with the approach being adopted here, is to be considered in terms of defining sequences. If, as asserted, its Fourier transform proves to be III(s), then it too will be considered in terms of sequences.

We proceed therefore to construct a sequence of regular Fourier transform pairs of ordinary functions such that one sequence is suitable for defining III(x), and we then see whether the other sequence defines III(s).

Consider the function

$$f(x) = \tau^{-1}e^{-\pi\tau^2x^2} \sum_{n=-\infty}^{\infty} e^{-\pi\tau^{-2}(x-n)^2}.$$

For a given small value of τ (which we shall later allow to vary to generate a sequence), the function $f(x)$ represents a row of narrow Gaussian spikes of width τ, the whole multiplied by a broad Gaussian envelope of width τ^{-1} (see Fig. 10.17). As $\tau \to 0$, each spike narrows in on an integral value of x and increases in height. For any value of x not equal to an integer, we can show that $f(x) \to 0$ as $\tau \to 0$ and, in addition, the area under each spike approaches unity. The sequence is therefore a suitable one for defining a set of equal unit impulses situated at integral values of x.

The function $f(x)$ possesses a regular Fourier transform, since $|f(x)|$ is integrable, and there are no discontinuities. The Fourier transform $F(s)$ is given by

$$F(s) = \tau^{-1} \sum_{m=-\infty}^{\infty} e^{-\pi\tau^2m^2}e^{-\pi\tau^{-2}(s-m)^2}.$$

One way of establishing this is to note that the factor

$$\tau^{-1} \sum_{n=-\infty}^{\infty} e^{-\pi\tau^{-2}(x-n)^2}$$

is periodic, with unit period, and hence may be expressed as a Fourier series. The theory of Fourier series is a well-established branch of mathematics that we are entitled to draw upon here, as long as we do not attempt to make the self-

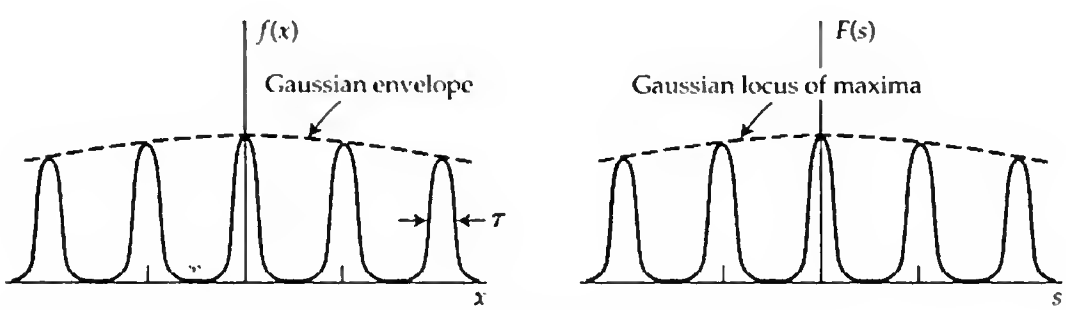


Fig. 10.17 **A transform pair for discussing the *shah* symbol.**

transforming property of the *shah* symbol a basis for reestablishing the theory of Fourier series. The Fourier series is

$$\sum_{m=-\infty}^{\infty} e^{-\pi\tau^2 m^2} \cos 2\pi m x.$$

Therefore

$$\begin{aligned} f(x) &= \sum_{m=-\infty}^{\infty} e^{-\pi\tau^2 m^2} e^{-\pi\tau^2 x^2} \cos 2\pi m x \\ &= \sum_{m=-\infty}^{\infty} e^{-\pi\tau^2 m^2} e^{-\pi\tau^2 x^2} e^{2\pi i m x}. \end{aligned}$$

By applying the shift theorem term by term, we obtain $F(s)$ in the form quoted above.

The function $F(s)$ is a row of Gaussian spikes of width τ with maxima lying on a broad Gaussian curve of width τ^{-1}. As before, it may be verified that as $\tau \to 0$ a suitable defining sequence for the *shah* symbol results.

BIBLIOGRAPHY

Blackman, R. B., and J. W. Tukey: "The Measurement of Power Spectra from the Point of View of Communications Engineering," Dover Publications, New York, 1958.

Jenkins, G. M., and D. G. Watts: "Spectral Analysis and Its Applications," Holden-Day, San Francisco, 1969.

Linden, D. A.: "A Discussion of Sampling Theorems," *Proc. IRE*, vol. 47, p. 1219, 1959.

Shannon, C. E.: "A Mathematical Theory of Communications," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423 and 623–656, 1948.

———: "Communication in the Presence of Noise," *Proc. IRE*, vol. 37, pp. 10–21, 1949.

Wiener, N.: "The Extrapolation, Interpolation and Smoothing of Stationary Time Series with Engineering Applications," Wiley, New York, 1949.

PROBLEMS

1. Show that a periodic function $p(x)$ with unit period can always be expressed in the form $\text{III}(x) * f(x)$ in infinitely many ways, and relate this to the fact that infinitely many different functions can share the same infinite set of equidistant samples.

2. Express the periodic pulse train

$$\sum_{n=-\infty}^{\infty} \Pi[10(x-n)]$$

in the form $\text{III}(x) * f(x)$ in three distinct ways.

3. **Series coefficients.** Determine the Fourier series coefficients for the functions of period equal to unity which, in the interval $-\frac{1}{2} < x < \frac{1}{2}$, are defined as follows:
   (*a*) $\cos \pi x$, $\Lambda(2x)$, $\Pi(2x) - \frac{1}{2}$,
   (*b*) $\Lambda(8x - 1)$, $(1 - 4x)H(x)$,
   (*c*) $e^{-\pi x^2}$, $1 - 4x^2$, $e^{-|x|}$, $e^{-x}H(x)$.

4. **Interpolation.** The following sample set defines a band-limited function:

$$\ldots\ 0, 10, 30, 50, 50, -40, -35, -10, -5, 0,\ \ldots$$

   All samples omitted are zero. Establish the form of the function by numerical interpolation. What is the minimum value assumed by the function?

5. **Undersampling.** A certain function is approximately band-limited, that is, a small fraction μ of its power spectrum in fact lies beyond the nominal cutoff frequency. It is sampled at the nominal sampling interval and reconstituted by the usual rules. Use the inequality

$$f(x) \leqslant \int_{-\infty}^{\infty} |F(s)|\, ds$$

   to examine how great the discrepancy between the original and reconstituted functions can become.

6. **Undersampling ordinate and slope.** The approximately band-limited function mentioned above is subjected to ordinate-and-slope sampling. Use the inequality

$$f'(x) \leqslant 2\pi \int_{-\infty}^{\infty} |sF(s)|\, ds$$

   to show that the discrepancy between the original and reconstituted functions can be serious, even when μ is small. ▷

7. **Sampling in the presence of noise.** A little noise is added to a band-limited function. Before sampling, one subjects the noisy function to filtering that eliminates the noise components beyond the original cutoff. However, the function is still contaminated with a little band-limited noise. Examine the relative susceptibilities of ordinate-and-slope sampling and of ordinary ordinate sampling to the presence of the noise.

8. **Self-convolution.** A finite sequence of equispaced impulses of finite strength

$$a_0\, \delta(x) + a_1\, \delta(x - 1) + \ \ldots\ + a_n\, \delta(x - n)$$

   is convolved with itself many times, and the result is another sequence of impulses

$$\sum \alpha_i\, \delta(x - i).$$

   Combine the central-limit theorem with the sampling theorem to show that a graph of α_i against i will approach a normal distribution. State the simple condition that the coefficients $a_0, a_1, \ldots, a_n$ must satisfy.

9. **Self-convolution.** In the previous problem, practice with simple cases that violate the condition, and develop the theory that suggests itself.

10. Gibbs overshoot. Show that the overshoot quoted as around 9 percent in the discussion of the Gibbs phenomenon is given exactly by

$$-\int_1^\infty \operatorname{sinc} x \, dx.$$

11. Sampling a bandpass signal. A carrier telephony channel extends from a low-frequency limit $f_l = 95$ kilohertz to a high-frequency limit $f_h = 105$ kilohertz. The signal is sampled at a rate of 21 kilohertz, and a new waveform is generated consisting of very brief pulses at the sampling instants with strength proportional to the sample values. Make a graph showing the spectral bands occupied by the new waveform, and verify that the original waveform could be reconstructed. Show that in general the critical sampling rate is $2f_h$ divided by the largest integer not exceeding $f_h/(f_h - f_l)$.

12. Interlaced sampling. In the previous problem, show that $2(f_h - f_l)$ is in general too slow a sampling rate to suffice to reconstitute a carrier signal. Show that by suitably interlacing two trains of equispaced samples the average sampling rate can, however, be brought down to twice the bandwidth.

13. Analogue filtering of data samples. A band-limited signal $X(t)$ passes through a filter whose impulse response is $I(t)$. The input and output signals $X(t)$ and $Y(t)$ can be represented by sample values X_i and Y_i. Show that

$$\{Y_i\} = \{I_i\} * \{X_i\}$$

and explain how to derive the sequence $\{I_i\}$.

14. Input from output. In the previous problem the input signal samples are $\{X_i\} = \{1\ 2\ 3\ 4\ 5\ \dots\}$, and the sequence $\{I_i\}$ describing the filter is $\{1\ 2\ 1\}$. Show that the output sequence $\{Y_i\}$ is $\{1\ 4\ 8\ 12\ \dots\}$ and that it is possible to work back from the known output and determine the input by evaluating

$$\{1\ \ -2\ \ 3\ \ -4\ \ 5\ \dots\} * \{1\ 4\ 8\ 12\ \dots\}.$$

15. Prediction by recursion. In the previous problem, verify numerically that a particular output signal sample can be expressed in terms of the history of the *output* plus a knowledge of the most recent input signal sample, that is,

$$Y_i = \alpha_i Y_{i-1} + \alpha_2 Y_{i-2} + \dots + \beta X_i.$$

Show that the coefficients $\alpha_1, \alpha_2, \dots$ and β are given in terms of the reciprocal sequence

$$\{K_0\ K_1\ K_2 \dots\} = \{I_i\}^{-1}$$

by
$$\{\alpha_1\ \alpha_2\ \alpha_3 \dots\} = -\frac{1}{K_0}\{K_1\ K_2\ K_3 \dots\}$$

and
$$\beta = \frac{1}{K_0}.$$

16. Bandpass filter. The input signal to a filter ceases, but the output continues, with each new sample deducible from the previous ones by the relation

$$Y_i = \alpha_1 Y_{i-1} + \alpha_2 Y_{i-2} + \dots .$$

Show that

$$Y(t) = [\alpha_1\,\delta(t-1) + \alpha_2\,\delta(t-2) + \ldots] * Y(t).$$

In a particular case, there are only two nonzero coefficients:

$$Y_i = 1.65Y_{i-1} - 0.9Y_{i-2}.$$

The first two output samples immediately after the input ceased were each 100; calculate and graph enough subsequent output values to determine the general character of the behavior. Show that a damped oscillation $Y(t) = e^{-\sigma t}\cos[\omega(t-a)]$ satisfies the convolution relation given above when

$$\alpha_1 = 2e^{-\sigma}\cos\omega$$

and

$$\alpha_2 = -e^{-2\sigma}$$

Is this band-limited behavior?

17. **General filter.** In the previous problem, show that the series for Y_i contains only a finite number of terms if the filter is constructed of a finite number of inductors, capacitors, and resistors. Hence show that the output due to any input signal is deducible, for a filter whose internal construction is unknown, after a certain number of consecutive sample measurements have been made at its terminals. How would you know that enough samples had been taken? ▷

18. **Unraveling a black box.** The input voltage $X(t)$, and output voltage $Y(t)$ of an electrical system are sampled simultaneously at regular intervals with the following results.

$X(t)$	15	10	6	2	1	0	0	0	0	0
$Y(t)$	15	15	7.5	−2.75	−2.5					

Calculate the missing values of $Y(t)$, and also calculate what the output would be if, after some time had elapsed, $X(t)$ began to rise linearly.

19. **Shah symbol.** Show that

$$\sum_{n=-\infty}^{\infty} e^{-i2\pi ns} = \text{III}(s).$$

20. **Sum of samples proportional to integral.** A bandlimited signal of long but finite duration is sampled at regular intervals T and the sample values are summed. Show that, provided a certain condition is met, the sum will be the same if the same total number of samples are taken, but at intervals that alternate between $T - b$ and $T + b$ (interlaced samples).

21. **Computer graphics.** A computer printout presents a long string of numbers which have to be graphed. Smith plots the points by hand, and as they seem to lie on a reasonably smooth curve, he draws a smooth curve through them by hand. Johnson, on the other hand, has the numbers plotted automatically using a computer program that joins successive points by a straight line. Smith says, "I can see by the sharp corners in your mechanical-looking curve that you have introduced spurious high frequencies that have no basis in fact." Johnson says, "My curve is more reliable because no subjective judgment enters into it." Robinson says, "I have put the Smith and Johnson

curves through my Fourier analyzer, and it is true that Johnson has a small high-frequency content that is absent from Smith's, but it is practically undetectable. The main discrepancy is that Smith's spectrum is noticeably stronger than Johnson's except for the d-c and neighboring low frequencies. If anything, it seems to be Smith who has added extra components, but in the middle-frequency range." Lee says, "Maybe the Smith curve is best and the mechanical plotting method has subtracted components." The time has come to do better than give a qualitative opinion; what do *you* say?

22. **Update operator.** Complicated input waveforms are to be applied to a linear time-invariant system and the response is to be computed step by step from an equation of the form

$$\Delta V_2(t) = U\{V_1(t)\},$$

where U describes an operation to be carried out on the part of $V_1(t)$ that has been applied up to any particular moment. Then, after a step Δt in time, exactly the same operation will be carried out on the updated segment of $V_1(t)$ and a new increment in $V_2(t)$ will be computed, and so on. What is this operation U? In what way might this method of computation be superior to simply convolving $V_1(t)$ with the impulse response $I(t)$?

23. **Associativity of multiplication and convolution.** Two functions are to be convolved but first one of them is to be sampled. Investigate whether $[f(x)\text{III}(x)] * g(x) = f(x) * [\text{III}(x)g(x)]$.

24. **Parseval's theorem.** If $p(x)$ is real and periodic with period T, show that

$$\frac{1}{T}\int_{-T/2}^{T/2}[p(x)]^2\,dx = a_0^2 + \frac{1}{2}\sum_{n=1}^{\infty}(a_n^2 + b_n^2),$$

where a_n and b_n are the Fourier coefficients of the Fourier series for $p(x)$.

25. **Interpolation.** Values of a function $f(x)$ are given at all integral values of x.
    (*a*) Show that

$$f(x) = \sum_{n=-\infty}^{\infty} f(n)\,\text{sinc}\,(x-n)$$

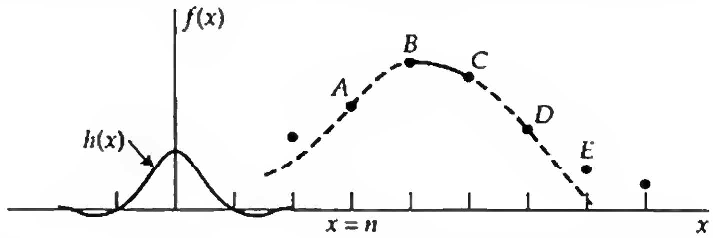
    is an interpolation formula that yields $f(x)$ correctly for any value of x, provided that $f(x)$ contains no frequencies as high as 0.5 cycle per unit of x or higher.
    (*b*) Show that this interpolation formula is expressible as a convolution between sinc x and $\sum_{n=-\infty}^{\infty} f(n)\delta(x-n)$.

26. **Spline interpolation.** Let $f(n)$, $f(n+a)$, $f(n+2a)$, and $f(n+3a)$ be four consecutive data values (labeled A, B, C, and D). It is proposed to interpolate as follows. A cubic curve (shown broken) will be put through A, B, C, and D (a cubic can be put through six points) and two extra conditions will be imposed. The slope at B will be the same as the slope of the line AC and the slope at C will be the same as the slope of the line BD. The arc BC will be accepted as the interpolation; then an arc CD will be determined in the same way using the points B, C, D, E, and so on. This is a form of spline interpolation, so called by analogy with the flexible splines used by draftsmen as a guide in drawing curved lines. Show that the operation described is expressible as convolution

with

$$h(x) = \begin{cases} 0.5a^{-3}(3|x|^3 - 5a|x|^2 + 2a^3) & 0 \le |x| \le a \\ 0.5a^{-3}(-|x|^3 + 5a|x|^2 - 8a^2|x| + 4a^3) & a \le |x| \le 2a \\ 0 & |x| \ge 2a \end{cases}$$

and work out the corresponding transfer function. Obtain the sequence of coefficients for midpoint interpolation and explain the value obtained for their sum.



27. Lagrange interpolation. Values of a function $f(x)$ are given for all integral values of x and interpolated values at $x = n + p$, where $0 < p < 1$, are calculated from the Lagrange four-point formula.

$$f(n + p) \approx -p(p - 1)(p - 2)\frac{f(n - 1)}{6} + (p^2 - 1)(p - 2)\frac{f(n)}{2}$$
$$-p(p + 1)(p - 2)\frac{f(n + 1)}{2} + p(p^2 - 1)\frac{f(n + 2)}{6}.$$

The interpolated value is a linear function of the given values, but can it be expressed in the form of a convolution, in which case Lagrange interpolation could be characterized as a filtering operation with a distinctive transfer function?

28. Finite number of samples. By measuring the flow in a river at a suitable site on different dates it is possible to establish a calibration curve relating the flow in cubic meters per second to the height of the water at a single height gage. After that the height can be measured automatically and telemetered to a central point where calculations can be made regarding future water storage in the dam receiving the discharge and future hydroelectric power production. In a certain network where the measurements from each site were made once a day, the predicted height of the dam based on the first 6 months of operation came out all wrong. Computer programmer Lee said, "Summing the daily flows and multiplying by 24 hours, which is what my program did, can only be an approximation to the actual volume delivered in 6 months, which is, strictly speaking, an integral, not a sum." Mathematician Long said, "Even if the height of a river is a bandlimited function of time, and even though we use 180 samples, there are an infinite number of functions, all with the same band limit, that pass through the 180 sample points. The key virtue of the sampling theorem is to permit interpolation, but the *interpolation formula requires an infinite number of samples*. We implicitly assumed that sample values prior to the first were zero. I have a proof that the set of functions, all passing through the N sample points, do not even have to be approximately the same, but may differ from each other by more than $|M|$, for any M no matter how large." Give your opinion on each sentence of the quoted statements. If you were the hydrologist, what would you have said?

29. **Sampling rate versus interpolation quality.** A data stream arrives in the form of a voltage which contains very little at frequencies above 1 Hz but enough to cause concern over aliasing. A careful investigation reveals that $S(f) = \exp(-\pi f^2)$ is as good a representation of the voltage spectrum as any. A single pulse of the form $V_0(t) = \exp(-\pi t^2)$ has this spectrum, so it is used for test purposes. A sampling device looks at $V_0(t)$ for 1 millisecond every 0.5 second. The samples are subsequently used to reconstruct a waveform by linear interpolation for display on an oscilloscope. The resulting polygon is only a rough approximation to the waveform $V_0(t)$. Over the interval from $t = -1$ to $t = +1$, what is the mean absolute error? Does the phasing of the samples relative to the time of occurrence of peak signal make a difference? If sinc function interpolation was used, more elaborate arrangements would be needed but what would the error be reduced to? If linear interpolation were retained but the sampling rate were doubled, what would the error be? What would you recommend?

30. **Aliasing.** It is necessary to predict the sunspot number 6 months ahead for scheduling frequencies for overseas radio communication. Each day the sunspot number is determined, as it has been for the last two centuries; at the end of each month the list is published, together with the mean value for the month. The monthly means, when graphed, give a rather jagged curve, not suited to the prediction of trends, so for each month a 12-month weighted running mean R_0' is calculated from the formula

$$R_0' = \frac{R_{-6}}{24} + \frac{R_{-5}}{12} + \frac{R_{-4}}{12} + \frac{R_{-3}}{12} + \frac{R_{-2}}{12} + \frac{R_{-1}}{12} + \frac{R_0}{12} + \frac{R_1}{12}$$
$$+ \frac{R_2}{12} + \frac{R_3}{12} + \frac{R_4}{12} + \frac{R_5}{12} + \frac{R_6}{24}$$

Naturally, this quantity is only available after a delay. From Fourier analysis of R_0', it is found that although R_0' is reasonably smooth, it nevertheless tends to have wiggles in it with a period of about 8 months; in fact, the Fourier transform peaks up at a frequency of 1/8.4 cycle per month.

Natural periodicities connected with the sun include the 11-year cycle and the 27-day interval between times when the sun presents roughly the same face to the earth, so it is hard to see how 8-month wiggles would have a solar origin. Can you explain why the wiggles are there?

31. **Theorem for band-limited functions.** An article in a current technical journal was being summarized before a discussion group by Smith, who said, "The author refers to a property of band-limited signals according to which two consecutive maxima cannot be more closely spaced than αT, where T is the critical sampling interval. Does anyone know this theorem?" Lee said, "It's more or less obvious. Consecutive maxima of a sinc function range from a greatest separation equal to 2.49 T down to $2T$ in the limit. A superposition of such sinc functions, which after all is what a band-limited function is, cannot possess maxima closer than $2T$. Therefore, $\alpha = 2$." Yanko, who had been thinking, then pointed out that $\sin 2t - \exp[-(t/1000)^2]$ would be adequately sampled at intervals $T = 1$ and made a sketch to show that the function possessed two rather closely spaced zeros at A and B. "I am pretty sure," he said, "that the integral of a band-limited function is also band-limited. Therefore, the maxima occurring at A and B in the integral of my function are spaced more closely than $2T$. What's more, I can make them

as close as I please by adjusting the amplitude of the sine wave." Confirm or disprove the various statements. What is your value of α?

32. **Harmonics.** The rails of an electric train system deliver d-c power obtained by 12-phase rectification of 60-hertz power to the locomotive, and the signals that are depended upon for safety precautions, such as keeping a maximum of one train in any one block, are delivered along the same rails but at audio frequency. (Signals are in a binary code generated by switching between two frequencies in the 5- to 10-kilohertz range, different pairs of frequencies being used in adjacent blocks.) The heavy currents have the same direction in the two rails and are nominally equal ("balanced"). The audio currents, on the other hand, are in opposite directions and are picked off the rails by two noncontacting ferrite-core coils whose outputs are combined so as to double the signal pickup and, in theory, cancel the ripple at power frequencies. As a final step, the signal frequencies are selected out by "maximally flat" quartz filters. In spite of this, the signaling system is not reliable enough to depend on for automatic train control.

The term "power frequencies" was used advisedly because the d-c motors draw ripple currents at 720 hertz (and harmonics) from the rails. To explain: let

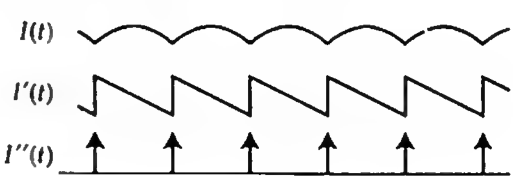
$$I_1(t) = \begin{cases} A \cos \omega_0 t & \text{when } \cos \omega_0 t > 0 \\ 0 & \text{when } \cos \omega_0 t < 0, \end{cases}$$

where $\omega_0 = 2\pi \times 60$ Hz and let

$$I(t) = I_1(t) + I_1(t - \tfrac{1}{720}) + I_1(t - \tfrac{2}{720}) + \ \dots \ + I_1(t - \tfrac{11}{720}).$$

which is almost pure d-c as a result of the combining of the 12 staggered waveforms but in fact has a fluctuating component with a fundamental frequency of 720 hertz. If, as happens when the train starts, the d-c component of $I(t)$ is 10,000 amperes, calculate the number of amperes flowing in each harmonic that falls in the range 5 to 10 kilohertz.

Consultant A makes wise remarks about the frequency stability of the audio oscillators, bit rate, signal bandwidth, filter bandwidth, and out-of-band rejection level but as an interim measure recommends inspecting and replacing all quartz filters that do not comply with specifications, and wire-brushing the rails. Consultant B states that the waveform $I(t)$ exhibits a scalloped form as illustrated and that the waveform picked up by the ferrite coils has a sawtooth form because a coil responds to the derivative of the flux through it. To suppress this sawtooth interference component in the desired audio signal, he proposed to differentiate once again; then the interference component will be strictly localized in time, in the form of brief pulses which may be clipped at a level that does not affect the desired sinusoidal signals. Is this right?



33. **Convolution of band-limited functions.** The acoustic energy flux incident on an airport bus stop as an aircraft takes off is to be measured with precision. The plan is to

digitally record samples of the air pressure $p(t)$ and velocity $v(t)$. Preliminary investigation reveals that both waveforms are band-limited and adequately sampled when the sampling interval is one unit of t. Then the energy flux per unit area is given by $\int_{-\infty}^{\infty} p(t)v(t)\,dt$. This procedure was adopted in the hope of avoiding the uncertainty of basing an energy measurement on pressure alone, especially in the presence of a simultaneous rush of hot fumes. A young systems-theory student on a summer job pointed out that the theoretically correct integral cannot be obtained from discrete data and offered to write a program for getting a least-mean-squares estimate of the integral from the discrete data. The engineer on the job said he always just took the products of the data values and added them up. "As a matter of fact," he said, "when I used to do this by hand I found that if you throw away every second value on both channels you only have half the work and you still get exactly the right answer." What is the truth of the matter?

34. **Fourier series.** Euler published the formula $x/2 = \sin x - \frac{1}{2}\sin 2x + \frac{1}{3}\sin 3x - \ldots$ well before Fourier's work on trigonometric series. Is it correct? What is there about this formula that causes Euler not to receive the credit for Fourier's series?

35. **Binomial coefficients.** The rth binomial coefficient $B(r, n)$ for a given n is given by $n!/r!(n-r)!$. Write a MATLAB M-file whose first line reads function [y] = binom(n) so that the string of $n+1$ coefficients for a stated n can be generated. For example, typing binom(5) should produce 1 5 10 10 5 1.

36. **Improved notation for stretched *shah* function.** When $\text{III}(x)$ is stretched by an integer factor L to give $\text{III}(x/L)$, the impulses that were originally at unit spacing are now spaced by L units; but they are no longer unit impulses: they have strength L. Mountararat says, "I propose an improved notation $\text{III}_L(x)$ to stand for unit impulses at spacing L. In the old notation my function would be represented as $(1/L)\text{III}(x/L)$." Would it be right to suppose that, in the improved notation, the Fourier transform of $\text{III}_L(x)$ would be $(1/L)\text{III}_{1/L}(s)$? ▷

37. **Inverse sampling theorem.** A signal waveform has a duration T; consequently the signal spectrum is fully determined by equispaced values separated by a frequency interval T^{-1}. Is it necessary for one of the frequency values to be zero? For example, can the area under the signal be determined from spectral samples at frequencies that do not include $f = 0$? ▷

38. **A Fourier series.** Show that $\sin x$ can be expanded as a cosine series as follows:

$$\frac{1}{2}\pi \sin x = 1 - \frac{2}{1.3}\cos 2x - \frac{2}{3.5}\cos 4x - \frac{2}{5.7}\cos 6x - \ldots, \qquad 0 \le x \le \pi$$

and explain how an odd function can be expanded as a sum of even functions. ▷

39. **Cotangent as a replication.** See whether the cotangent function can be represented as a superposition of $1/x$ functions suitably displaced along the x-axis so that their poles coincide with the poles of $\cot x$ at $x = 0, \pm\pi, 2\pi, \ldots$. In other words, investigate

$$\cot x = \sum_{k=-\infty}^{\infty} \frac{1}{x - k\pi}. \triangleright$$

40. Checking analysis by computer. It has been suggested that $\text{III}(x)\ \text{sgn}\ x$ has Fourier transform $-i \cot \pi s$. The innermost pair of impulses $-\delta(x+1)+\delta(x-1)$ has FT $-2i \sin 2\pi s$; then transforming pair by pair suggests that

$$\sum_{k=1}^{\infty} 2 \sin 2\pi k s = \cot \pi s.$$

To check for a sign error, a missing factor of 2, or more serious errors, compute both sides for a single value of s; for example, does $\sin 1^\circ + \sin 2^\circ + \sin 3^\circ \ldots$ add up to $\frac{1}{2} \cot \frac{1}{2}^\circ$? Graph the sum of N terms versus N and discuss your finding. ▷

11

The Discrete Fourier Transform and the FFT

If one wishes to obtain the Fourier transform of a given function, it may happen that the function is defined in terms of a continuous independent variable, as is most often the case in books, especially in lists of transform pairs. But it may also happen that function values are given only at discrete values of the independent variable, as with physical measurements made at regular time intervals. Regardless of the form of the given function, if the transform is evaluated by numerical computing, the values of the transform will be available only at discrete intervals. We often think of this as though an underlying function of a continuous variable really exists and we are approximating it. From an operational viewpoint, however, it is irrelevant to talk about the existence of values other than those given and those computed (the input and output). Therefore, it is desirable to have a mathematical theory of the actual quantities manipulated.

THE DISCRETE TRANSFORM FORMULA

Questions of discreteness also arise in connection with periodic functions. A periodic function is describable by a sequence of coefficients at discrete intervals (of frequency), but this situation *may* be viewed as a special case of continuous frequency. The transform is then regarded as a string of equally spaced delta functions (Fig. 11.1) of strengths given by the coefficients. What if this transform were itself periodic? Then the original function would also reduce to a string of delta functions. With both function and transform now being periodic, all the information about both would be limited to two finite sets of coefficients: the strengths of the delta functions. Periodic impulse trains were discussed earlier, towards the end of Chapter 10.

Thus the practical situation where a finite set of values is given and a finite set is computed does actually lie within the continuous theory. However, it pays

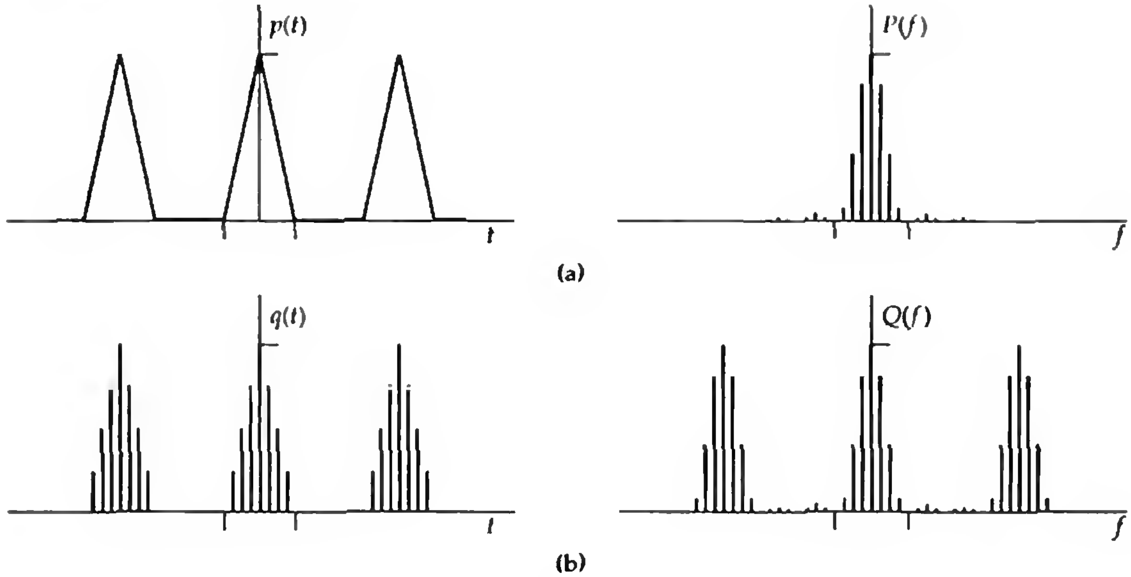


Fig 11.1 (a) A periodic function $p(t)$ has a transform $P(f)$ which is a string of equispaced delta functions; (b) a periodic function $q(t)$, which is a string of equispaced delta functions, has a transform $Q(f)$ which is also a string of equispaced delta functions and the string is periodic.

to start afresh rather than to force the theory of the discrete Fourier transform into the continuous framework. The reason is that the discrete notation is concise and reasonably standardized.

To retain some physical ties, let us think in terms of a signal that is a function of time; but to recognize the discreteness of the independent variable, let us use the symbol τ, which we agree can assume only a finite number N of consecutive integral values. Furthermore, we agree that τ cannot be negative. Thus, before entering into the realm of the discrete Fourier transform, we first make, if necessary, a change of scale and a change of origin. For example, suppose that a voltage waveform $v(t)$ is half a period of a cosinusoid of period 1 second (Fig. 11.2):

$$v(t) = \begin{cases} \cos 2\pi t & -0.25 < t < 0.25 \\ 0 & \text{otherwise} \end{cases}$$

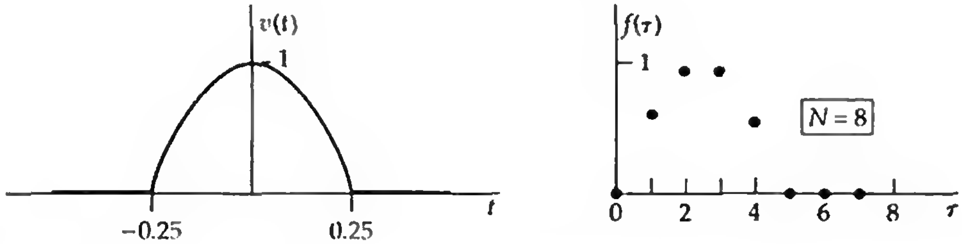


Fig. 11.2 A function of the continuous variable t and one way of representing it by eight sample values.

TABLE 11.1

t (milliseconds)	$v(t)$
−250	0
−150	0.588
−50	0.951
50	0.951
150	0.588
250	0

TABLE 11.2

τ	$f(\tau)$
0	0
1	0.588
2	0.951
3	0.951
4	0.588
5	0

and that samples are taken at intervals of 100 milliseconds. Table 11.1 shows signal values as a function of t but Table 11.2 shows how it is to be converted into a function $f(\tau)$ of discrete time τ before proceeding. In general, if the sampling interval is T and the first sample of interest occurs at $t = t_0$, then

$$f(\tau) = v(t_0 + \tau T), \qquad \tau = 0, 1, 2, \ldots, N - 1.$$

In what follows, $f(\tau)$ forms the point of departure. It will be noticed that no provision is made for cases where there is no starting point, as with a function such as $\exp(-t^2)$. This is in keeping with the practical character of the discrete transform, which does not contemplate data trains dating back to the indefinitely remote past. A second feature to note is that the finishing point must occur after a finite time. However, it need not come at $\tau = 5$ as in Table 11.2; one might choose to let τ run on to 15 and assign values of zero to the extra samples. This is a conscious choice that must always be made. It may be important; for example, Table 11.2 does not convey the information given in the equation preceding it—that following the half-period cosine, the voltage remains zero. The table remains silent on that point, and if it is important, the necessary number of zeros would need to be appended.

By *definition*, $f(\tau)$ possesses a discrete Fourier transform $F(\nu)$ given by

$$F(\nu) = N^{-1} \sum_{\tau=0}^{N-1} f(\tau) e^{-i2\pi(\nu/N)\tau}. \tag{1}$$

Just as the leading coefficient a_0 in the Fourier series expansion of a periodic function equals the mean of the function values, so the leading value $F(0)$ of the discrete Fourier transform equals the mean $N^{-1}\sum f(\tau)$ of the values of $f(\tau)$. Respect for prior convention is the reason for the factor N^{-1} in the definition. In computing practice, as distinct from formal definition, it is efficient to combine the N^{-1} with later normalizing factors or graphical scale factors than to prematurely multiply every element of $F(\nu)$ by N^{-1}.

The quantity ν/N is analogous to frequency measured in cycles per sampling interval. The correspondence of symbols may be summarized as follows:

	Time	Frequency
Continuous case	t	f
Discrete case	τ	ν/N

The symbol ν has been chosen in the discrete case, instead of f, to emphasize that the frequency integer ν is *related* to frequency but is not the *same* as frequency f. For example, if the sampling interval is 1 second and there are eight samples ($N = 8$), then the component of frequency f will be found at $\nu = 8f$; conversely, the frequency represented by a frequency integer $\nu = 1$ will be $\frac{1}{8}$ hertz.

Given the discrete transform $F(\nu)$, one may recover the time series $f(\tau)$ with the aid of the inverse relationship

$$f(\tau) = \sum_{\nu=0}^{N-1} F(\nu)e^{i2\pi(\nu/N)\tau}. \tag{2}$$

To see that this is so, we first verify the fact that

$$\sum_{\nu=0}^{N-1} e^{-i2\pi(\nu/N)(\tau-\tau')} = \begin{cases} N & \tau = \tau' \\ 0 & \text{otherwise.} \end{cases}$$

One way is to picture the summation as a closed polygon on the complex plane, except when $\tau = \tau'$, in which case the polygon becomes a straight line composed of N unit vectors end to end. (The variable τ' is taken to assume values 0, 1, ..., $N - 1$; if larger values were allowed, for example if $\tau - \tau'$ could become equal to N, $2N$, ..., the summation would equal N for $\tau = \tau'$ mod N.) Another way to think of this is by analogy with the orthogonality relation of sinusoids of different frequencies, a viewpoint that is aided by rewriting it

$$\sum_{\tau=0}^{N-1} e^{-i2\pi(\nu/N)\tau}e^{i2\pi(\nu'/N)\tau} = \begin{cases} N & \nu = \nu' \\ 0 & \text{otherwise.} \end{cases}$$

To establish the inverse discrete transform, introduce a dummy variable τ' for convenience, and substitute (1) into the right-hand side of (2):

$$\begin{aligned}
\sum_{\nu=0}^{N-1} F(\nu)e^{i2\pi(\nu/N)\tau'} &= \sum_{\nu=0}^{N-1} N^{-1} \sum_{\tau=0}^{N-1} f(\tau)e^{-i2\pi(\nu/N)\tau}e^{i2\pi(\nu/N)\tau'} \\
&= N^{-1} \sum_{\tau=0}^{N-1} f(\tau) \sum_{\nu=0}^{N-1} e^{-i2\pi(\nu/N)(\tau-\tau')} \\
&= N^{-1} \sum_{\tau=0}^{N-1} f(\tau) \times \begin{cases} N & \tau = \tau' \\ 0 & \tau \neq \tau' \end{cases} \\
&= f(\tau').
\end{aligned}$$

The definition of the inverse differs by having a positive sign in the exponent and in having no preceding factor N^{-1}.

From the inverse transform we see that only N integral values of the frequency integer ν are needed and that they range from 0 to $N - 1$ just as with the discrete time τ. It certainly sounds reasonable that a function defined by N measurements should be representable after transformation by just N parameters. Even when the values of f are real, the values of F are in general complex and, as will be seen below, one must be careful how to count complex numbers. The fact that ν ranges over integer values, whereas the frequency ν/N is fractional, is the reason for introducing the integer ν; mathematical convenience takes priority over physical significance.

In order to regain our physical feeling for numerical orders of magnitude, let us consider a record consisting of 1,024 samples separated by 1-second intervals. We expect this to be representable by a Fourier series consisting of a constant term and multiples of a certain fundamental frequency. The fundamental period should be 1,024 seconds, corresponding to a fundamental frequency 1/1,024 hertz. The highest frequency needed will be 0.5 hertz, which has two samples per period. This will be the 512th harmonic. The reason that ν assumes 1,024 values, whereas the number of frequencies is only 512, is as follows. If the values of $f(\tau)$ are real, as is usual in records of physical quantities, there are 1,024 real data values. Now the transform $F(\nu)$ has 1,024 complex values which would require 2,048 real numbers to specify except that $F(0)$ and $F(N/2)$ have no imaginary part [see (Eq. 1)] and half the remaining values of $F(\nu)$ are complex conjugates of the other half. This is because $f(\tau)$ is real (Chapter 2). If $f(\tau)$ were complex, there would be 2,048 real data values and $F(\nu)$ would require 2,048 real numbers for its specification.

Since the highest frequency reached is 0.5 hertz, it is apparent that ν/N, which reaches a maximum value of 1,023/1,024, is not a *strict* analogue of frequency. For instance, where $\nu = 724$, $\nu/N = 0.707$ hertz and, as we have seen, frequencies above 0.5 hertz are neither required nor can be represented by samples at 1-second intervals. Rather, $\nu = 724$ corresponds to a negative frequency $-300/1{,}024$ hertz.

This anomaly is a distinct impediment to the visualization of the connection between the Fourier transform and the discrete Fourier transform. One way to bring the two into harmony would be to redefine the discrete transform in terms of summation over negative and positive values, although it might be objected that negative subscripts are not permitted in some computer languages. Figure 11.3*b* shows how a signal $v(t)$ and its transform $S(f)$ might be made to correspond with $f(\tau)$ and its discrete transform $F(\nu)$. Alternatively, as in Fig. 11.3*c*, it could be arranged that $\tau = 0$ corresponds to $t = t_0$; this would not prevent the small negative values of frequency labeled A from corresponding to large positive values of ν labeled B.

Finally, another way of looking at this, which is customary, is to lift the restriction of τ and ν to values from 0 to $N - 1$ and to allow all integral values, including negative values. The function values assigned are on the basis that $f(\tau)$ and $F(\nu)$, in their extended sense, are periodic, with period N. Thus

$$f(\tau) = f(\tau \pm N) = f(\tau \pm 2N) = \dots$$

$$F(\nu) = F(\nu \pm N) = F(\nu \pm 2N) = \dots .$$

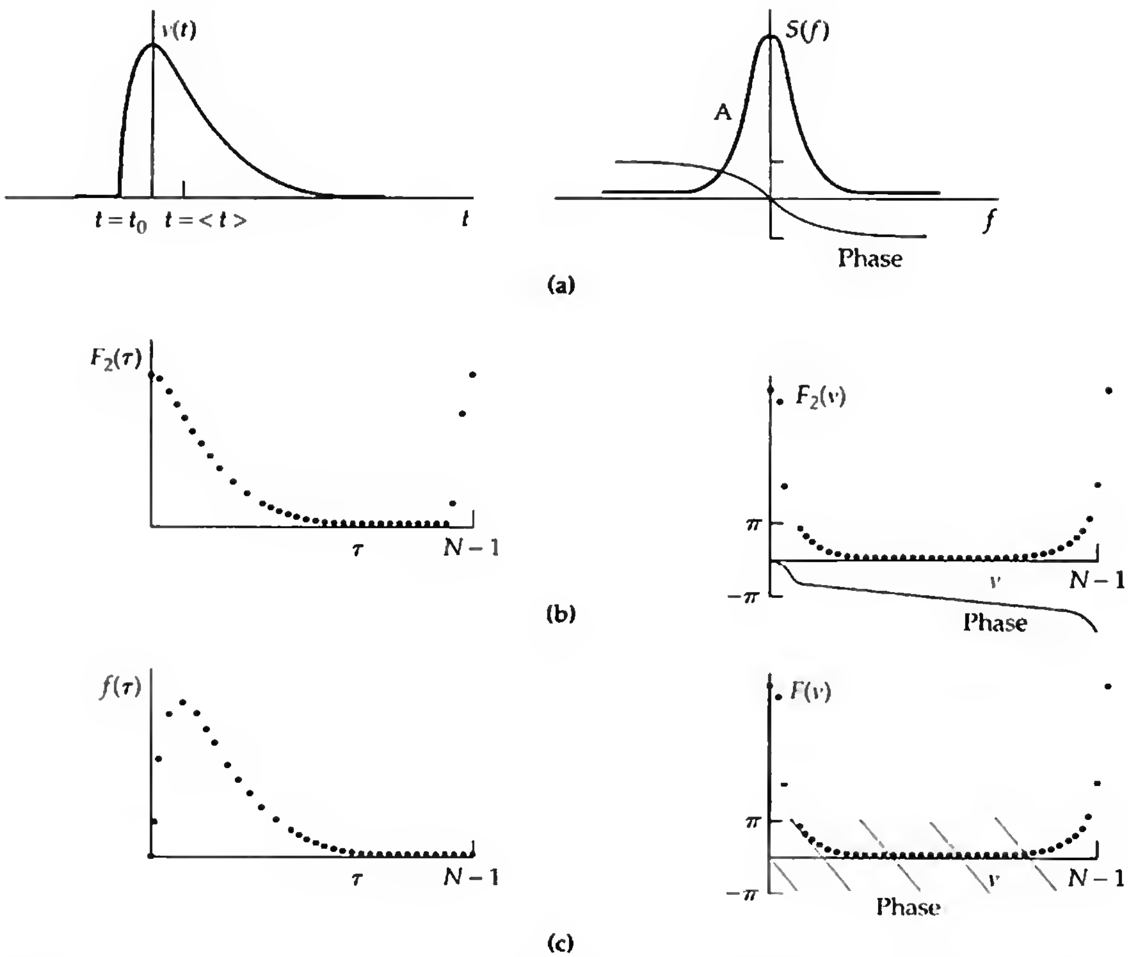


Fig. 11.3 **(a) A function and its Fourier transform; (b) and (c) two ways of representing the function by N samples and the corresponding discrete Fourier transforms, shown by modulus (dots) and phase (small dots).**

This plan will be adopted here. Under this plan, the basic transform and its inverse may be written

$$F(\nu) = N^{-1} \sum_{\tau=-N/2}^{N/2-1} f(\tau) e^{-i2\pi(\nu/N)\tau} \tag{3}$$

$$f(\tau) = \sum_{\nu=-N/2}^{N/2-1} F(\nu) e^{i2\pi(\nu/N)\tau} \tag{4}$$

and ν/N may be identified with frequency measured in cycles per sampling interval over the range $-\frac{1}{2}N \leq \nu \leq \frac{1}{2}N$. If the sampling interval is T, the frequency measured in hertz is ν/NT.

In Fig. 11.4, we see a way of visualizing $f(\tau)$ and $F(\nu)$ as having cyclic dependence on τ and ν. The upper type of diagram is helpful in making a connection with our previous experience [e.g., to take the autocorrelation of a sequence,

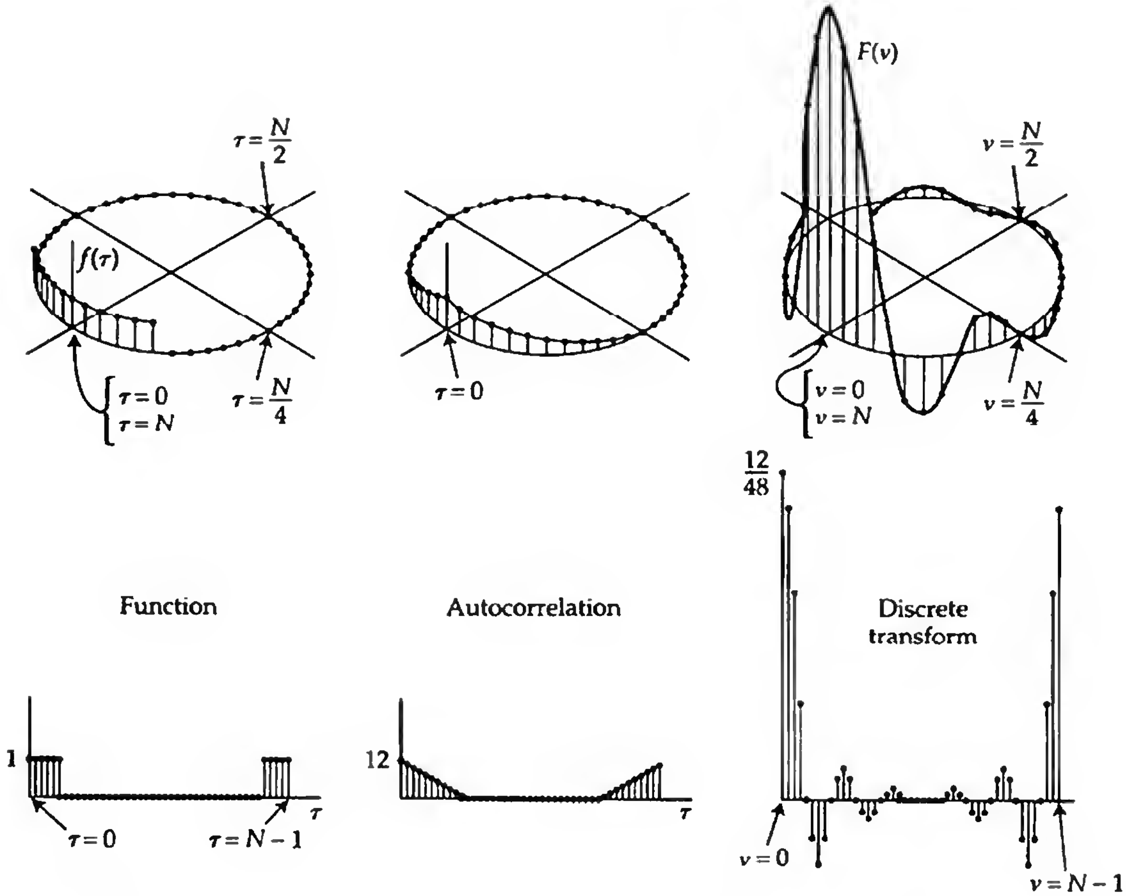


Fig. 11.4 Cyclic (above) and standard (below) representations of a sampled rectangle function, its triangular autocorrelation, and its transform.

we imagine $f(\tau)$ to be rotated, in the top left-hand diagram, relative to itself]. Multiplication of corresponding values followed by summation gives a value of the autocorrelation, and we can see how the result ties in with our previous knowledge that the autocorrelation of a rectangle function is a triangle function. This illuminates the lower diagrams where the original form of indexing, starting from zero, obscures the simple shapes.

CYCLIC CONVOLUTION

If we convolve two sequences, one having m elements and the other n, then the convolution sum, or serial product, will have $m + n - 1$ elements (p. 32). In particular, if we deal with sequences having N elements, then the convolution of two such sequences will not itself be containable in an N-element sequence. Figure

11.4 (top center) coped with the desire to exhibit the triangular autocorrelation of two rectangles by packing the sequence expressing the rectangle with extra zeros. Three-fourths of the elements shown at the top left are zeros, and at the top center there are still plenty of zeros left to witness to the isolated nature of the triangular island.

Suppose now that N is kept constant while the number of nonzero elements at the top left is increased. When the rectangle extends more than halfway around the circle, the outer ends of the triangle overlap. The result will be a flat, nonzero segment in the region of overlap. Clearly, the result is wrong if we are looking for the triangular autocorrelation. However, the operation exists in its own right, and may be defined as cyclic convolution.

The cyclic convolution integral $h(\theta)$ of two periodic functions $f(\cdot)$ and $g(\cdot)$ with period 2π is defined as

$$h(\theta) = \int_0^{2\pi} f(\theta')g(\theta - \theta')\, d\theta'.$$

The cyclic convolution sum $h(\tau)$ of two N-element sequences $f(\tau)$ and $g(\tau)$ is defined as

$$h(\tau) = \sum_{\tau'=0}^{N-1} f(\tau')g[\tau - \tau' + NH(\tau' - \tau)],$$

where $H(\cdot)$ is the Heaviside unit step function. Because $\tau - \tau'$ may range from $-(N-1)$ to $N-1$ as τ and τ' range from 0 to $N-1$, a term N must be added to $\tau - \tau'$ when $\tau' > \tau$ to bring $\tau - \tau'$ into the range $[0, N-1]$. This is the effect of the term $NH(\tau' - \tau)$, which will be explicitly required in computations. However, interpreting $g(\tau)$ in its extended or cyclic sense allows us to omit the term $NH(\tau' - \tau)$ from the written expressions.

EXAMPLES OF DISCRETE FOURIER TRANSFORMS[1]

Certain short, discrete transform pairs are often needed. A small stock of transform pairs can be helpful for checking and is recommended when studying new algorithms and transforms. Here is a reference list for the DFT.

$N = 2$:

$$\begin{aligned} \{1\ 0\} &\supset \tfrac{1}{2}\{1\ 1\} \\ \{1\ 1\} &\supset \tfrac{1}{2}\{2\ 0\} \\ \{0\ 1\} &\supset \tfrac{1}{2}\{1\ -1\} \\ \{1\ -1\} &\supset \tfrac{1}{2}\{0\ 2\}. \end{aligned}$$

[1] As there is little risk of confusion, we may use the sign $\supset$ to stand for "has DFT." Thus Eq. (1) could be written $f(\tau) \supset F(\nu)$ and Eq. (2) could be written $F(\nu) \subset f(\tau)$, where $\subset$ means "has inverse DFT."

$N = 4$:

$$
\begin{aligned}
\{1\ \ 0\ \ 0\ \ 0\} &\supset \tfrac{1}{4}\{1\ \ 1\ \ 1\ \ 1\}\\
\{0\ \ 1\ \ 0\ \ 0\} &\supset \tfrac{1}{4}\{1\ \ -i\ \ -1\ \ i\}\\
\{0\ \ 0\ \ 1\ \ 0\} &\supset \tfrac{1}{4}\{1\ \ -1\ \ 1\ \ -1\}\\
\{0\ \ 0\ \ 0\ \ 1\} &\supset \tfrac{1}{4}\{1\ \ i\ \ -1\ \ -i\}\\
\{1\ \ 1\ \ 0\ \ 0\} &\supset \tfrac{1}{4}\{2\ \ 1-i\ \ 0\ \ 1+i\}\\
\{0\ \ 0\ \ 1\ \ 1\} &\supset \tfrac{1}{4}\{2\ \ -1+i\ \ 0\ \ -1-i\}\\
\{1\ \ 1\ \ 1\ \ 1\} &\supset \tfrac{1}{4}\{4\ \ 0\ \ 0\ \ 0\}\\
\{1\ \ 1\ \ 0\ \ -1\} &\supset \tfrac{1}{4}\{1\ \ 1-2i\ \ 1\ \ 1+2i\}\\
\{1\ \ 0\ \ 0\ \ 1\} &\supset \tfrac{1}{4}\{2\ \ 1+i\ \ 0\ \ 1-i\}.
\end{aligned}
$$

$N = 8$:

$$
\begin{aligned}
\{1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\} &\supset \tfrac{1}{8}\{8\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0\}\\
\{1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\} &\supset \tfrac{1}{8}\{1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\}\\
\{0\ 1\ 0\ 0\ 0\ 0\ 0\ 0\} &\supset \tfrac{1}{8}\{1\ \ e^{-i2\pi/8}\ \ e^{-i2\pi(2/8)}\ \ e^{-i2\pi(3/8)}\\
&\qquad -1\ \ e^{-i2\pi(5/8)}\ \ e^{-i2\pi(6/8)}\ \ e^{-i2\pi(7/8)}\}\\
\{0\ 0\ 1\ 0\ 0\ 0\ 0\ 0\} &\supset \tfrac{1}{8}\{1\ \ -i\ \ -1\ \ i\ \ 1\ \ -i\ \ -1\ \ i\}.
\end{aligned}
$$

RECIPROCAL PROPERTY

Just as the Fourier transformation applied twice in succession, with alternating sign of i, gives back the original function, so there is a corresponding property for the discrete Fourier transform. But owing to the use of a scaling factor N to make the index ν integral, the discrete transform is not strictly reciprocal, even allowing for the sign of i. Thus

$$N^{-1}\left[\sum_{\nu=0}^{N-1} N^{-1}\sum_{\tau=0}^{N-1} f(\tau')e^{-i2\pi(\nu/N)\tau'}\right]e^{i2\pi(\nu/N)\tau} = N^{-1}f(\tau).$$

If the DFT is applied twice in succession without changing the sign of i, we will get $N^{-1}f(-\tau)$ on the right-hand side. Expressing this differently, if

$$f(\tau) \supset F(\nu),$$

then

$$F(\nu) \supset N^{-1}f(-\tau).$$

This property and other properties, and several theorems will not be derived, but are readily verifiable by numerical trial. Instead, emphasis will be placed on interpreting and illustrating them and presenting them in a form suitable for reference.

ODDNESS AND EVENNESS

By definition, $f(\tau)$ is even if

$$f(-\tau) = f(\tau)$$

and odd if

$$f(-\tau) = -f(\tau).$$

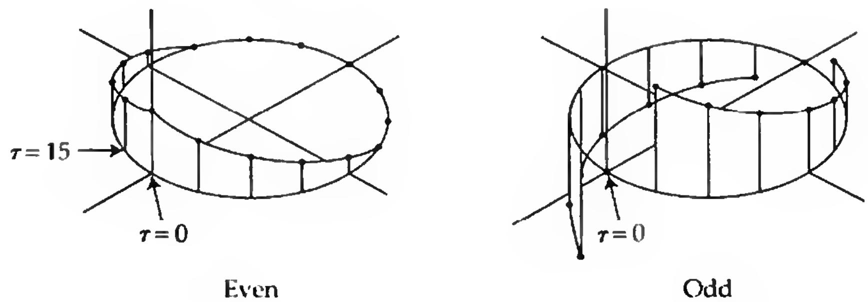


Fig. 11.5 **Even and odd sequences shown cyclically.**

Figure 11.5 illustrates the following odd and even sequences of 16 elements, where τ runs from 0 to 15.

$$\begin{array}{ll} \text{Even:} & \{5\,4\,3\,2\quad 1\,0\,0\,0\quad 0\,0\,0\,0\quad 1\,2\,3\,4\}. \\ \text{Odd:} & \{0\,8\,7\,6\quad 5\,4\,3\,2\quad 0\,-2\,-3\,-4\quad -5\,-6\,-7\,-8\}. \end{array}$$

In these examples, the elements have been grouped in fours to help display the nature of the symmetry. Even and odd sequences of four elements are as follows:

$$\begin{array}{ll} \text{Even:} & \{a\,b\,c\,b\}. \\ \text{Odd:} & \{0\,b\,0\,-b\}. \end{array}$$

Clearly the rules for extending the range of τ to negative integers are required here, in particular the special case

$$f(-\tau) = f(N - \tau).$$

A similar relation holds for $F(\nu)$,

$$F(-\nu) = F(N - \nu).$$

EXAMPLES WITH SPECIAL SYMMETRY

Symmetry rules with respect to oddness and evenness assume the same form for the discrete transform as given in Chapter 2 for the continuous transform.

$$\begin{array}{rcl} \text{real} & \supset & \text{hermitian} \\ \text{imaginary} & \supset & \text{antihermitian} \\ \text{real and even} & \supset & \text{real and even} \\ \text{real and odd} & \supset & \text{imaginary and odd} \\ \text{imaginary and even} & \supset & \text{imaginary and even} \\ \text{imaginary and odd} & \supset & \text{real and odd} \\ \text{even} & \supset & \text{even} \\ \text{odd} & \supset & \text{odd.} \end{array}$$

The following examples illustrate these results, which are of major importance in practical computing

$$\begin{aligned}
\{1\ 2\ 3\ 4\} &\supset \tfrac{1}{4}\{10\quad -2+2i\quad -2\quad -2-2i\}\\
i\{1\ 2\ 3\ 4\} &\supset \tfrac{1}{4}\{10i\quad -2-2i\quad -2\quad 2-2i\}\\
\{1\ 0\ 0\ 0\} &\supset \tfrac{1}{4}\{1\ 1\ 1\ 1\}\\
\{0\ 1\ 0\ -1\} &\supset \tfrac{1}{4}\{0\ -2i\ 0\ 2i\}\\
i\{4\ 2\ 1\ 2\} &\supset \tfrac{1}{4}i\{9\ 3\ 1\ 3\}\\
i\{0\ 1\ 0\ -1\} &\supset \tfrac{1}{4}\{0\ 2\ 0\ -2\}\\
\{1+4i\quad 2i\ i\ 2i\} &\supset \tfrac{1}{4}\{1+9i\quad 1+3i\quad 1+i\quad 1+3i\}\\
\{0\quad 1+i\quad 0\quad -1-i\} &\supset \tfrac{1}{4}\{0\quad 2-2i\quad 0\quad -2-2i\}.
\end{aligned}$$

COMPLEX CONJUGATES

The discrete transform of the conjugate is the conjugate of the transform, reversed:

$$f^*(\tau) \supset F^*(-\nu).$$

By "reversal," we mean changing the sign of the independent variable.

REVERSAL PROPERTY

If the sign of τ is changed, that is, if $f(\tau)$ is reflected in the line $\tau = 0$, the sign of ν is changed:

$$f(-\tau) \supset F(-\nu).$$

It is worth noting that this operation on $f(\tau)$ is also produced by reflection in the line $\tau = \frac{1}{2}N$.

ADDITION THEOREM

$$f_1(\tau) + f_2(\tau) \supset F_1(\nu) + F_2(\nu).$$

Example:

If	$\{2\ 0\ 0\ 0\} \supset \frac{1}{4}\{2\ 2\ 2\ 2\}$
and	$\{0\ 1\ 0\ 0\} \supset \frac{1}{4}\{1\ -i\ -1\ i\}$,
then	$\{2\ 1\ 0\ 0\} \supset \frac{1}{4}\{3\quad 2-i\quad 1\quad 2+i\}$.

SHIFT THEOREM

$$f(\tau - T) \supset e^{-i2\pi T(\nu/N)}F(\nu).$$

Example:

$$\begin{aligned}
\{1\,0\,0\,0\} &\supset \tfrac{1}{4}\{1\ 1\ 1\ 1\}\\
\{0\,1\,0\,0\} &\supset \tfrac{1}{4}\{1\ \ -i\ \ (-i)^2\ \ (-i)^3\}\\
\{0\,0\,1\,0\} &\supset \tfrac{1}{4}\{1\ \ (-i)^2\ \ (-i)^4\ \ (-i)^6\}\\
\{0\,0\,0\,1\} &\supset \tfrac{1}{4}\{1\ \ (-i)^3\ \ (-i)^6\ \ (-i)^9\}.
\end{aligned}$$

Sometimes a shift in the frequency domain is required; one version of what could be called the inverse shift theorem is

$$e^{i2\pi(\nu_0/N)\tau} f(\tau) \supset F(\nu - \nu_0).$$

CONVOLUTION THEOREM

The cyclic convolution of two sequences $\{f_1(\tau)\}$ and $\{f_2(\tau)\}$ was defined by

$$f_1(\tau) * f_2(\tau) = \sum_{\tau'=0}^{N-1} f_1(\tau') f_2(\tau - \tau').$$

Remember that $f_2(\cdot)$ has to be understood in its extended cyclic sense. To emphasize the distinction between this discrete operation and the convolution *integral*, we may use the term convolution *sum*, but where there is no risk of confusion, it may be called simply the convolution of the sequences $\{f_1\}$ and $\{f_2\}$. The theorem is

$$f_1(\tau) * f_2(\tau) \supset N F_1(\nu) F_2(\nu).$$

Example:

Let $\{1\ \ 1\ \ 0\ \ 0\} \supset \tfrac{1}{4}\{2\ \ \ 1-i\ \ \ 0\ \ \ 1+i\}.$

Then $\{1\ \ 1\ \ 0\ \ 0\} * \{1\ \ 1\ \ 0\ \ 0\} = \{1\ \ 2\ \ 1\ \ 0\} \supset \tfrac{1}{4}\{4\ \ \ -2i\ \ \ 0\ \ \ 2i\}.$

Also, $\{1\ \ 1\ \ 0\ \ 0\} * \{0\ \ 0\ \ 2\ \ 2\} = \{2\ \ 0\ \ 2\ \ 4\} \supset \tfrac{1}{4}\{8\ \ \ 4i\ \ \ 0\ \ \ -4i\}.$

PRODUCT THEOREM

The inverse of the convolution theorem, which applies to products in the τ domain, or convolution in the ν domain, is

$$f_1(\tau) f_2(\tau) \supset \sum_{\nu'=0}^{N-1} F_1(\nu') F_2(\nu - \nu').$$

Example:

$$f_1 = f_2 = f_1 f_2 = \{1\ \ 1\ \ 0\ \ 0\} \supset \tfrac{1}{4}\{2\ \ \ 1-i\ \ \ 0\ \ \ 1+i\}.$$

CROSS-CORRELATION

$$\sum_{\tau'=0}^{N-1} f_1(\tau')\, f_2(\tau' + \tau) \supset NF_1(\nu)F_2(-\nu).$$

AUTOCORRELATION

$$\sum_{\tau'=0}^{N-1} f_1(\tau')f_1(\tau' + \tau) \supset N|F_1(\nu)|^2.$$

Example:

$$\begin{aligned}\{1\ 1\ 0\ 0\} \star \{1\ 1\ 0\ 0\} &= \{2\ 1\ 0\ 1\}\\ &\supset \tfrac{1}{4}\{4\ 2\ 0\ 2\}.\end{aligned}$$

SUM OF SEQUENCE

$$\sum_{\tau=0}^{N-1} f(\tau) = NF(0).$$

Example:

$$\{1\,0\,5\,0\} \supset \{1.5\ \ -1\ \ 1.5\ \ -1\}.$$

We see that $\Sigma f = 6$, $F(0) = 1.5$, $N = 4$, and $NF(0) = 6$. If we follow the practice of writing N^{-1} as a first factor of $F(\nu)$, the theorem means that the sum of the sequence is equal to the first term after the opening brace on the right-hand side.

FIRST VALUE

This is the inverse of the preceding:

$$f(0) = \sum_{\nu=0}^{N-1} F(\nu).$$

Example:

$$f(0) = 1 \qquad \text{and} \qquad \Sigma F = 1.5 - 1 + 1.5 - 1 = 1.$$

We see that the mean of a sequence equals the first value of its DFT, but conversely the first value of the sequence equals the sum of the DFT.

mean equals A

$$\{a \ldots\} \supset \{A \ldots\}$$

first value a equals sum

GENERALIZED PARSEVAL-RAYLEIGH THEOREM

$$\sum_{\tau=0}^{N-1} |f(\tau)|^2 = N \sum_{\nu=0}^{N-1} |F(\nu)|^2.$$

Example:

$$\{1\ 1\ 0\ 0\} \supset \tfrac{1}{4}\{2 \quad 1-i \quad 0 \quad 1+i\}.$$

We see that $\Sigma f^2 = 2$ and that $N\Sigma F^2 = 4 \times 0.5 = 2$.

PACKING THEOREM

The packing operator Pack_K packs a given N-member sequence $f(\tau)$ with trailing zeros so as to increase the number of elements to KN.

$$\mathrm{Pack}_K\{f(\tau)\} = \{g(\tau)\},$$

where

$$g(\tau) = \begin{cases} f(\tau) & 0 \leqslant \tau \leqslant N-1 \\ 0 & N \leqslant \tau \leqslant KN-1. \end{cases}$$

Thus

$$\mathrm{Pack}_2\{1\,2\,3\,4\} = \{1\,2\,3\,4\,0\,0\,0\,0\}.$$

This theorem is

$$\mathrm{Pack}_K\{f(\tau)\} \supset G(\nu),$$

where $G(\nu) = \dfrac{1}{K} F\left(\dfrac{\nu}{K}\right)$, $\quad \nu = 0, K, 2K, \ldots, KN-K.$

The intermediate values of $G(\nu)$, not given by this relation, can be determined by sinc-function interpolation between the known values [e.g., by midpoint interpolation (p. 225) when $K = 2$], but for a better method, see Problem 11.8.

SIMILARITY THEOREM

To have an analogy with expansion or contraction of the scale of continuous time, we must supply sufficient zero elements, either at the end, as with packing, so that the sequence may expand, or between elements, so that there is room for contraction. The operation of inserting zeros between elements so as to increase the total number of elements by a factor K will be denoted by the stretch operator Stretch_K.

$$\text{Stretch}_K\,\{f(\tau)\} = \{g(\tau)\},$$

where
$$g(\tau) = \begin{cases} f(\tau/K) & \tau = 0, K, 2K, \ldots, (N-1)K \\ 0 & \text{otherwise.} \end{cases}$$

Example:

$$\text{Stretch}_2\,\{1\;2\;3\;4\} = \{1\;0\;2\;0\;3\;0\;4\;0\}.$$

The theorem is, if $\{g\} \supset \{G\}$,

$$G(\nu) = \begin{cases} \dfrac{1}{K}F(\nu) & \nu = 0, \ldots, N-1 \\ \dfrac{1}{K}F(\nu - N) & \nu = N, \ldots, 2N-1 \\ \cdots\cdots\cdots\cdots & \cdots\cdots\cdots\cdots\cdots\cdots \\ \dfrac{1}{K}F(\nu - \overline{K-1}N) & \nu = (K-1)N, \ldots, KN-1. \end{cases}$$

Thus, stretching by a factor K in the τ domain results in K-fold repetition of $F(\nu)$ in the ν domain; the frequency scale is not compressed by a factor K.

EXAMPLES USING MATLAB

The DFT is readily evaluated using MATLAB[R], a high-level language that can execute operations on arrays from the keyboard and therefore is particularly adapted to displaying numerical results as the sequence of complex values constituting the DFT of a given sequence. For example, to obtain the DFT of the sequence {1 1 1 1 0 0 0 0} just type

```
fft([1 1 1 1 0 0 0 0]),
```

press the return key, and the following is displayed:

4.0000 1.0000−2.4142*i* 0 1.0000−0.4142*i* 0 1.0000+0.4142*i* 0 1.0000+2.4142*i*

The MATLAB definition of the operator **fft**( ) does not include division by N, consequently the first value 4.0000 shows the sum of the input values rather than their mean. The DFT definition in this chapter includes division by N and so ensures that first value of the DFT sequence is the mean of the input values, as has long been established for the first coefficient a_0 of a Fourier series. Further minor differences vis-à-vis the DFT are seen from the definitions. Calling the input sequence $x(n)$, $1 \leqslant n \leqslant N$, the computed discrete transform values are defined by

$$X(k) = \sum_{n=1}^{N} x(n)e^{-i2\pi(k-1)(n-1)/N},$$

while the slightly different inverse summation that yields the original $x(n)$ is

$$x(n) = \frac{1}{N}\sum_{k=1}^{N} X(k)e^{i2\pi(k-1)(n-1)/N},$$

where the suppressed factor[2] $1/N$ now raises its head. The inverse summation is performed by **ifft**( ).

The variable n is the serial number of the input samples. For example, the N elements of a time series are to be indexed so that the first value is at $n = 1$ and the last value is at $n = N$. The connection between the variable k and frequency can be explained as follows. If the unit of n is one second, then when $k \leqslant N/2 + 1$ the frequency in hertz is $(k-1)/N$. This is not the standard[3] use of k in digital signal processing, where k is directly proportional to frequency (for $k \leqslant N/2$); still, the leftmost value of k in the display corresponds to zero frequency, just what one is accustomed to.

From continuous-variable theory it is known that applying the Fourier transform to a function $f(t)$ twice in succession produces $f(-t)$, the reverse of the original function. To test MATLAB against expectation, type

fft(**fft**([1 1 1 1 0 0 0 0]))

and obtain

8.0000 0.0000 + 0.0000i 0 + 0.0000i −0.0000 + 0.0000i,
0 8.0000 − 0.0000i 8.0000 − 0.0000i 8.0000 − 0.0000i

---

[2] In practical computing as distinct from algebraic definition, where consistency with previous practice is advised, it is wasteful to include the factor $1/N$ in the evaluation of either the direct or inverse summation. The stage following transformation usually calls for an additional multiplier for normalization to unit area or to unity at the origin or for adjusting the height of graphics output. Combinations of numerical factors can be reduced to a single multiplication at the final stage.

[3] It is customary to expect that $a_k \cos 2\pi kt$ represents the kth harmonic and in particular that $k = 0$ refers to the d.c. component at zero frequency, while $k = 1$ refers to the first harmonic (of unit frequency). With MATLAB, $k = 1$ refers to the d.c. component and $k = 2$ refers to the first harmonic.

which for ease of inspection may be rewritten

$$8\{1 \quad 0 \quad 0 \quad 0 \quad 0 \quad 1 \quad 1 \quad 1\}.$$

Apart from the factor 8 arising from the definitions, this is in fact the correct reverse of the original, under the same rules as for the DFT (see Fig. 11.3*b*). The leftmost element of n discrete samples of $f(t)$ corresponds to $t = 0$, the next element to the right corresponds to $t = \Delta t$, while the rightmost element corresponds to $t = -\Delta t$.

A symmetrical function of t, say $\Pi(t/5)$, sampled at intervals $\Delta t = 1$, would be represented by {1 1 1 0 0 0 1 1}. The DFT should be purely real since the given sequence is symmetrical about $n = 1$. Typing

fft([1 1 1 0 0 0 1 1])

yields

$$\{5 \quad 2.4142 \quad -1 \quad 0.4142 \quad -1 \quad 2.4142\}$$

which, we see, turns out to be purely real, in agreement with expectation for a symmetrical real input.

Exercise a. **Representation of an even function.** An even function of t has the property that $f(-t) = f(t)$ but a discrete sample set $x(n)$, where n runs from 1 to 8, does not allow for negative indices; therefore we cannot say that a sample set is even when $x(-n) = x(n)$. Satisfy yourself empirically that the sequence {4 2 1 0 0 0 1 2} is, however, suitable for representing samples of an even function on the grounds that its discrete Fourier transform is purely real. ▷

Exercise b. **Replacement rule for indices that are zero or negative.** Verify that the sequence {4 3 2 1 0 1 2 3} is suitable for representing the even function $4\Lambda(t/3.5)$ at unit intervals of t if the following replacement is made where n is nonpositive: $x(-n) \Rightarrow x(8 - n)$. Check that this replacement also accounts for the result of the previous exercise. ▷

Exercise c. **Applying the index replacement rule.** Consider a sequence {1 2 3 4 4 3 2 1 0}, which consists of samples taken at one second intervals of a function $f(t)$. Note that the values are samples of $\Lambda[(t-4)/4]$ (which is not an even function) taken at $t = 1, 2, \ldots 9$. First show that applying **fft( )** to the given sequence produces imaginary parts that are zero and that the result cannot therefore be even a rough approximation to the Fourier transform of $\Lambda[(t-4)/4]$, which will be complex. What then is an even function of t that the given sequence could refer to? ▷

Exercise d. **Odd functions.** Verify that the sequence {0 3 2 1 0 −1 −2 −3} is suitable for representing an odd function by showing empirically that operating on it with **fft( )** produces a purely imaginary discrete transform. Check that

the sampled function could be the odd function $\Lambda(t/4)\,\mathrm{sgn}\,t$. Would another possibility be $\Lambda(t/4)\,\mathrm{sgn}\,t + \sin \pi t$? ▷

***Exercise e.* Zero packing.** Let $x_4(n) = \{1\,1\,1\,1\}$, $x_8(n) = \{1\,1\,1\,1\,0\,0\,0\,0\}$, and $x_{16}(n) = \{1\,1\,1\,1\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\}$, all generated by sampling $\Pi[(t-2.5)/4]$ at intervals $\Delta t = 1$. Apply `fft(  )` to all three, compare the three results, and reconcile with the Fourier transform $F(s) = 4\exp(-i3\pi s)\,\mathrm{sinc}\,4s$. ▷

THE FAST FOURIER TRANSFORM

In 1965 a method of computing discrete Fourier transforms suddenly became widely known (Cooley and Tukey, 1965) and revolutionized many fields where onerous computing was an impediment to progress. Good sources of historical information are the IEEE Transactions on Audio and Electroacoustics, vol. AU-2, June 1967 and Bergland (1969). The name Discrete Fourier Transform was introduced in Good (1951).

There are various ways of understanding this fast Fourier transform (FFT). One way, which will appeal to certain people, is in terms of factorization of the transform matrix. From the definition, we can write the DFT relation (for $N = 8$) in the form of a matrix product,

$$\begin{bmatrix} F(0)\\ F(1)\\ F(2)\\ F(3)\\ F(4)\\ F(5)\\ F(6)\\ F(7) \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1\\ 1 & W & W^2 & W^3 & W^4 & W^5 & W^6 & W^7\\ 1 & W^2 & W^4 & W^6 & W^8 & W^{10} & W^{12} & W^{14}\\ 1 & W^3 & W^6 & W^9 & W^{12} & W^{15} & W^{18} & W^{21}\\ 1 & W^4 & W^8 & W^{12} & W^{16} & W^{20} & W^{24} & W^{28}\\ 1 & W^5 & W^{10} & W^{15} & W^{20} & W^{25} & W^{30} & W^{35}\\ 1 & W^6 & W^{12} & W^{18} & W^{24} & W^{30} & W^{36} & W^{42}\\ 1 & W^7 & W^{14} & W^{21} & W^{28} & W^{35} & W^{42} & W^{49} \end{bmatrix} \times \begin{bmatrix} f(0)\\ f(1)\\ f(2)\\ f(3)\\ f(4)\\ f(5)\\ f(6)\\ f(7) \end{bmatrix}, \qquad (1)$$

where $W = \exp(-i2\pi/N)$. The quantity W is an Nth root of unity, since $W^N = \exp(-i2\pi) = 1$. It may be thought of as a complex number whose modulus is unity and whose phase is $-(1/N)$ turns.

$$\begin{bmatrix} F(0)\\ F(1)\\ F(2)\\ F(3)\\ F(4)\\ F(5)\\ F(6)\\ F(7) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0 & 0 & W & 0 & 0\\ 0 & 0 & 1 & 0 & 0 & 0 & W^2 & 0\\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & W^3\\ 1 & 0 & 0 & 0 & W^4 & 0 & 0 & 0\\ 0 & 1 & 0 & 0 & 0 & W^5 & 0 & 0\\ 0 & 0 & 1 & 0 & 0 & 0 & W^6 & 0\\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & W^7 \end{bmatrix} \times \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0\\ 0 & 1 & 0 & W^2 & 0 & 0 & 0 & 0\\ 1 & 0 & W^4 & 0 & 0 & 0 & 0 & 0\\ 0 & 1 & 0 & W^6 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0\\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & W^2\\ 0 & 0 & 0 & 0 & 1 & 0 & W^4 & 0\\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & W^6 \end{bmatrix}$$

$$\times \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & W^4 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & W^4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & W^4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & W^4 \end{bmatrix} \times \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} f(0) \\ f(1) \\ f(2) \\ f(3) \\ f(4) \\ f(5) \\ f(6) \\ f(7) \end{bmatrix}. \qquad (2)$$

This factorization leaves only two nonzero elements in each row. In (1) there are N^2 multiplications but there are only $2N$ multiplications per factor if we use (2), and the number of factors M is given by $2^M = N$ if we do not count the first factor, which merely represents a rearrangement. Thus the multiplications total $2N \log_2 N$. Examination of the factors shows that many of the multiplications are trivial, and therefore to calculate the precise time saving will require careful attention to details. Nevertheless, we are better off by a factor of the order of $N/log_2 N$, which becomes very important for the large values of N which arise with very long data trains or with digitized two-dimensional images such as photographs, for example.

Here is another method of understanding the fast Fourier transform. A sequence of N elements may be divided into two shorter sequences of $N/2$ elements each by placing the even-numbered elements into the first sequence and the odd-numbered ones into the second. For example, $\{8\,7\,6\,5\,4\,3\,2\,1\}$ can be split into $\{8\,6\,4\,2\}$ and $\{7\,5\,3\,1\}$. Each of these possesses a DFT. From these two DFT's how could one obtain the DFT of the longer sequence? The answer is obtained by writing

$$\{8\,7\,6\,5\,4\,3\,2\,1\} = \{8\,0\,6\,0\,4\,0\,2\,0\} + \{0\,7\,0\,5\,0\,3\,0\,1\}.$$

We see that the desired DFT can be obtained by using the stretching and shift theorems. From the stretching theorem we know that if

$$\{8\,6\,4\,2\} \supset \{A\,B\,C\,D\},$$

then

$$\{8\,0\,6\,0\,4\,0\,2\,0\} \supset \tfrac{1}{2}\{A\,B\,C\,D\,A\,B\,C\,D\}, \qquad (3)$$

a phenomenon that may be familiar from Fourier series coefficients for periodic functions.

Likewise, if

$$\{7\,5\,3\,1\} \supset \{P\,Q\,R\,S\},$$

then

$$\{7\,0\,5\,0\,3\,0\,1\,0\} \supset \tfrac{1}{2}\{P\,Q\,R\,S\,P\,Q\,R\,S\}.$$

Now we apply the shift theorem to find that

$$\{0\,7\,0\,5\,0\,3\,0\,1\} \supset \tfrac{1}{2}\{P\;\; WQ\;\; W^2R\;\; W^3S\;\; W^4P\;\; W^5Q\;\; W^6R\;\; W^7S\}. \qquad (4)$$

Multiplication by W means rotation by one Nth of a revolution in the complex plane, so the effect of the shift is to apply a phase delay that increases progres-

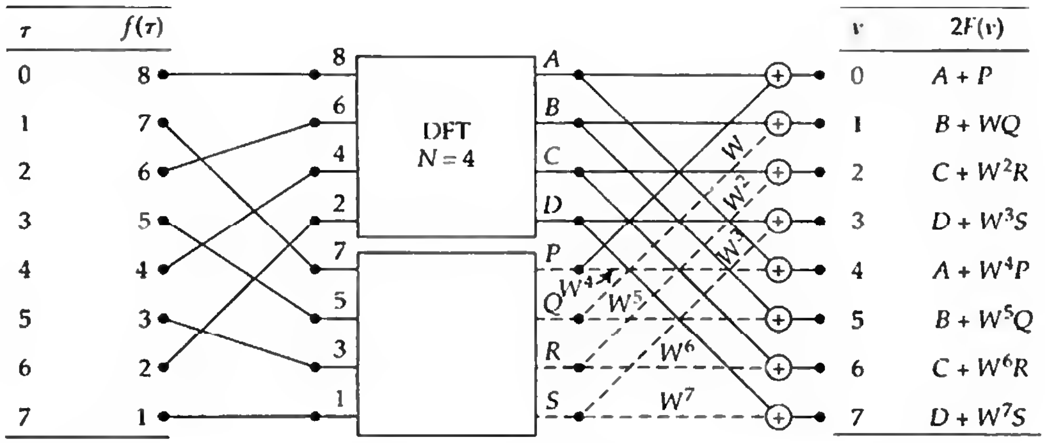


Fig. 11.6 **Reduction of eight-element DFT to two four-element DFTs.**

sively along the sequence of elements $\{P\,Q\,R\,S\,P\,Q\,R\,S\}$. Adding (3) and (4) gives the DFT of the long sequence. Thus the transformation with $N = 8$ has been broken down into two transformations with $N = 4$, which potentially represents a 50 percent time saving, since the number of multiplications in a DFT performed according to (1) goes as N^2. To see how this breaking down can be taken even further, we refer to Fig. 11.6. Starting with the given sequence on the left, we rearrange it into the two short sequences $\{8\,6\,4\,2\}$ and $\{7\,5\,3\,1\}$ that form the inputs to two transformers with $N = 4$ whose outputs are $\{A\,B\,C\,D\}$ and $\{P\,Q\,R\,S\}$, respectively. The unbroken flow lines show that A, B, C, and D are transferred to the output nodes to deliver $\{A\,B\,C\,D\,A\,B\,C\,D\}$. The broken flow lines are tagged with factors that cause the delivery of P, WQ, W^2R, etc., as in (4) to the same output nodes, where addition takes place. Figure 11.7 now shows a further reduction of each four-element transformer to two two-element transformers, and Fig. 11.8 shows the full reduction to multiplications and additions.

Finally, the steps may be summarized as follows. First, we rearrange the given sequence into $\{8\,4\,6\,2\,7\,3\,5\,1\}$, an operation corresponding exactly to multiplica-

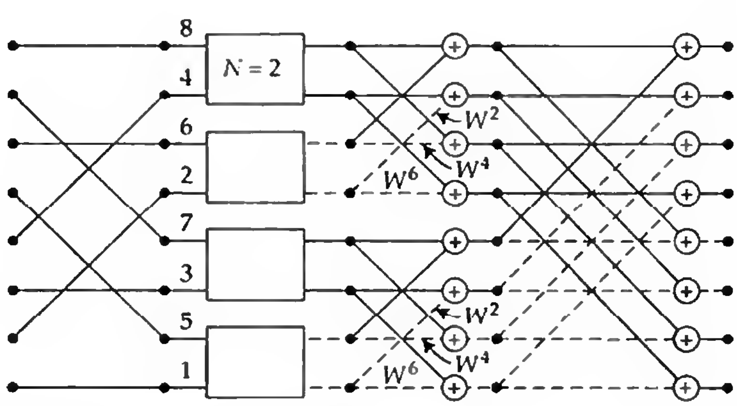


Fig. 11.7 **Reduction to four two-element DFTs.**

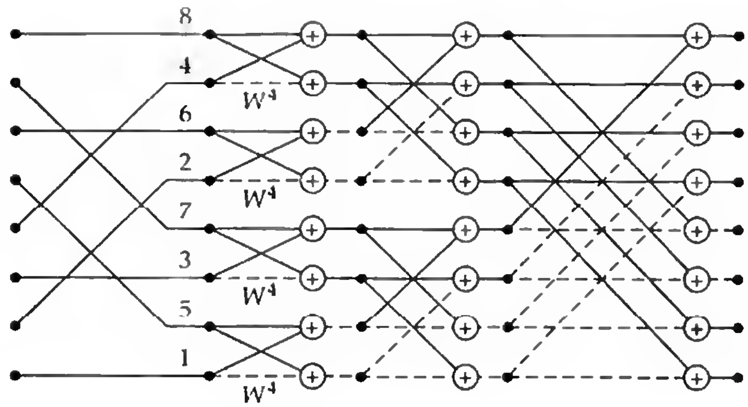


Fig. 11.8 **Reduction of eight-element DFT to 3 × 16 multiplications and 3 × 8 additions. In the foregoing three figures, multiplication by unity is shown by a full line and the broken lines are associated with the factors shown. (Adapted from Cochran et al., 1967.)**

tion by the first square matrix of (2) and sometimes loosely referred to as bit reversal. Then eight new numbers are calculated as linear combinations of various pairs of the rearranged data, exactly as indicated by the second square matrix of (2). These numbers are the outputs from the left-hand column of adders in Fig. 11.8. There are two more similar stages, making a total of three such operations in all (or M, in general, where $2^M = N$). Of course, not all the 48 multiplications are significant. There are 32 multiplications by unity and 7 multiplications by W^4, which is simply a sign reversal. In addition, W^2 and W^6 are rather simple to handle.

We thus see that Fig. 11.8 is an intimate representation of the four matrix multiplications of (2), and the considerations leading to the construction of Fig. 11.8 could be the basis of a step-by-step discovery of the factors that were presented in (2) without derivation.

If the number of elements cannot be halved indefinitely (i.e., N is not expressible as 2^M), a fast algorithm may still be tailormade to suit. The final reduction might then incorporate three-element transformers if, for example, N was divisible by 3. Such algorithms are not quite as fast.

PRACTICAL CONSIDERATIONS

Many practical considerations have been incorporated into the programs available for performing the fast Fourier transform. For some applications, speed is an overriding consideration; for others, convenience. If N is not a power of 2, convenience says pack the data with zeros; speed says choose a modified program taking advantage of such factors as N possesses. Some users never require complex output; others do, but not in the form of real and imaginary parts. Some re-

quire two and three dimensions. Some normally have to segment their data because N exceeds the capacity of their computer. Questions of this kind, though important, can be studied through the literature using the sources given above or by examination of existing programs. Examination of documentation can be particularly important, because some software packages do not implement the DFT at all but rather some modification possessing more or less convenience.

Here we shall look into the very common situation where a data string of, say, 60 elements is to be transformed by a general-purpose program which allows a choice of 64, 128, or other power of 2 for N. Adding four zeros will fit the data to the program. Will it make a difference whether the zeros trail, precede, or are placed in twos at start and finish? From Fig. 11.4 and remembering the shift theorem, we see that $|F(\nu)|$ will be unaffected but that the effective shifts of origin will introduce phase differences. If phase is important, as it might be if the data string has a natural origin, the shift theorem will supply the appropriate phase correction factor.

Will it make a difference if 68 trailing zeros are added and the $N = 128$ program is used? Surprisingly to some, the answer can be yes. To understand this, consider the extended sense in which $f(\tau)$ and $F(\nu)$ are regarded as periodic with period N.

Let $v(t)$ be a function of the continuous variable t, which at integral values of t in the range just embracing 0 to $N - 1$ agrees with $f(\tau)$ and is zero outside, as in Fig. 11.9*a*. Then, $v(t)\text{III}(t)$ is a string of impulses (Fig. 11.9*b*) that contains precisely the same information as $f(\tau)$ but does not have the property of repeating

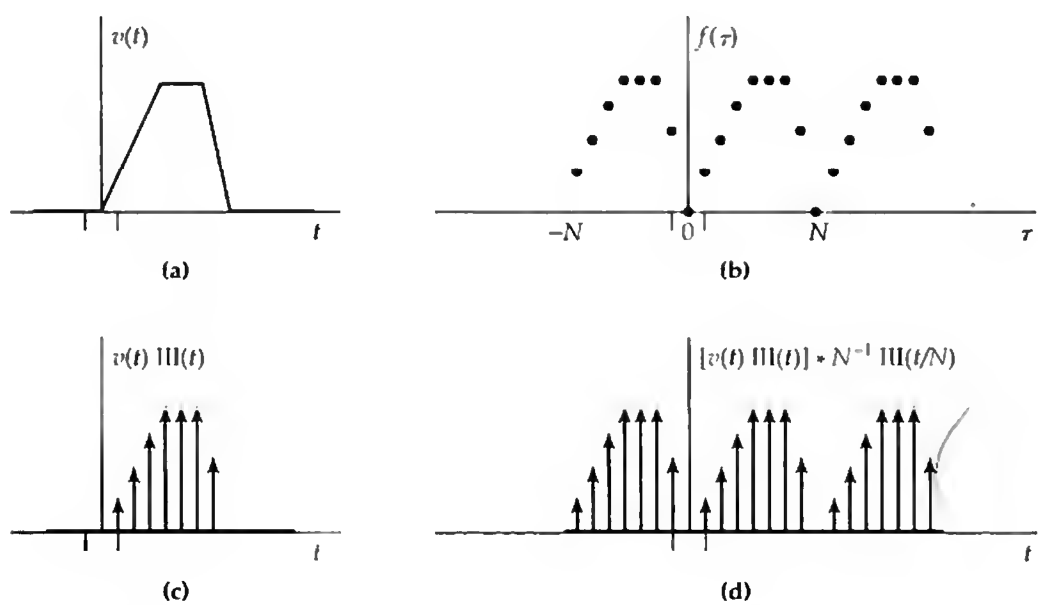


Fig. 11.9 **A function (a) and its discrete representation (b) on a cyclic basis. Parts (c) and (d) show two pulse waveforms, as functions of continuous time, that are equivalent to the discrete representation.**

with period N. Periodic character is, however, exhibited by the expression

$$p(t) = [v(t)\mathrm{III}(t)] * N^{-1}\mathrm{III}(t/N),$$

which is in strict analogy with $f(\tau)$. Because of the convention of representing impulses by arrows with length equal to the strength of the impulse, Fig. 11.9*c* would become a precise representation of $f(\tau)$ if the abscissa label were changed to τ and the arrowheads were changed to blobs.

If the Fourier transform of $v(t)$ is $S(f)$, then we may use the convolution theorem and knowledge that the *shah* function $\mathrm{III}(\cdot)$ transforms into itself to obtain

$$P(f) = [S(f) * \mathrm{III}(f)]\mathrm{III}(Nf).$$

This expression has the same exact correspondence with the DFT $F(\nu)$ that $p(t)$ has with $f(\tau)$, provided that we use the relation $f = \nu/N$ to translate between f and ν. The factor N^{-1} in (3) is accounted for. Thus the rather simple algebra developed in Chapter 10 includes as a special case discrete situations that appear at first to be outside the scope of the integral transform.

If we now replicate the same data string $v(t)\mathrm{III}(t)$ with a period $2N$, we have $[v(t)\mathrm{III}(t)] * (2N)^{-1}\mathrm{III}(t/2N)$ and the transform becomes

$$[S(f) * \mathrm{III}(f)]\mathrm{III}(2Nf).$$

The only difference is that the same expression $S(f) * \mathrm{III}(f)$ is sampled twice as closely in frequency. How, then, could an apparently improved sampling be perceived as different? The answer is that $S(f)$ may be oscillatory and indeed normally will be, unless $f(\tau)$ is free from large jumps such as those that often occur at the beginning and end of data strings. Of course, because of the cyclic character of $f(\tau)$, a large initial value $f(0)$ will not count as a large jump if the final value $f(N-1)$ is approximately equal. But if such a data string of 64 elements were extended to 128 elements by the addition of trailing zeros, there *would* be jumps. Rough structure would then appear in $F(\nu)$. Likewise, if four consecutive elements of the relatively smooth 64-element string were put to zero, oscillations would appear in $F(\nu)$. This suggests that adding zeros might not always be the best way to pack a data string out to 64 elements. A result more in keeping with expectation might result from packing with dummy values having more plausibility as data than zeros would have.

IS THE DISCRETE FOURIER TRANSFORM CORRECT?

While the theory of the DFT is precise and self-consistent and exactly describes the manipulations performed on actual data samples when a Fourier transform is to be computed, the question remains to what degree the DFT approximates the Fourier transform of the function underlying the data samples. Clearly, the DFT can be only an approximation, since it provides only for a finite set of discrete frequencies. But will these discrete values themselves be correct? It is easy

to show simple cases where they are not. Discussion of the question can be based on the sampling theorem and the phenomenon of aliasing. If the initial samples are not sufficiently closely spaced to represent high-frequency components present in the underlying function, then both the DFT values and a smooth curve passing through them will be falsified by aliasing. If the underlying function is known, then the error associated with a given choice of sampling interval is calculable. If one takes the operational viewpoint that the measured data samples may be the only knowledge we have, then avoidance of error will depend on experimental factors such as prior knowledge or experience. For example, a run with twice the number of samples in the same time could confirm the presence or absence of higher frequencies.

A further important source of error in the DFT lies in the truncation of data strings. It is, of course, unavoidable that truncation of a function will result in an erroneous Fourier transform (the result obtained will be the convolution of the true Fourier transform with a certain sinc function), and so truncation error is not specific to the DFT. However, the error committed will be different. To see this, imagine a case where the sampling interval is quite fine enough to cope with the highest frequencies present in the data, so that there is no aliasing error. Now truncate the data. The effect on the DFT will be to convolve it with samples of the sinc function corresponding to the width of the rectangle-function factor describing the truncation. But this time we are convolving with an entity such as $Q(f)$ in Fig. 11.1. In addition to the smoothing out we now have the prospect of the left and right islands of $Q(f)$ leaking into the central island. The truncation effect thus comprises both smoothing error, or reduction of fine detail in the DFT, and leakage error. Leakage error may be reduced, at the expense of increased smoothing error, by use of a tapered truncation factor in place of the rectangle-function factor. The best compromise must depend on the case; leakage error tends to falsify the "higher" frequencies (ν in the neighborhood of $N/2$), whereas smoothing error is distributed differently.

APPLICATIONS OF THE FFT

In some subjects, such as x-ray diffraction and radio interferometry, the observational data require Fourier transformation in order to be presented in customary ways, such as a molecular shape, a crystal structure, or a brightness distribution map of a celestial source. In these fields the introduction of the FFT merely speeds up what was already practiced.

In other applications, one takes the Fourier transform in order to perform some operation on it and then retransforms. For example, if we had a photographic enlargement that was particularly grainy (i.e., finely speckled because of the grain structure within the photographic emulsion) we might subject the photograph to two-dimensional low-pass filtering. First we would digitize it into a two-dimensional array of numbers, although it might already be in digitized form

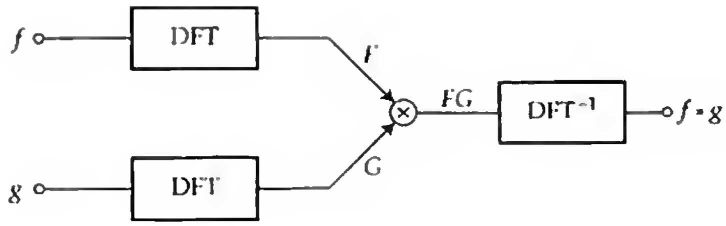


Fig. 11.10 **Flow diagram for convolving.**

(e.g., if it had been received by radio telemetry from a space probe). Then we would take the two-dimensional DFT and remove or reduce the higher spatial frequencies by multiplication with a suitable low-pass transfer function. Finally, we would invert the DFT. Of course this would be equivalent to convolving the digitized photograph with the appropriate point-source response (inverse DFT of the transfer function). For desk calculation one convolution may be more attractive than two DFT's and one set of multiplications. But with the large quantities of data that a photograph normally contains, a larger computer would be required, and it would then be found that the DFT route is quicker if the FFT is used. The reason is that, if there are N elements in the array of data, the number of multiplications required is of the order of N^2, whereas as we have seen, the FFT requires far fewer if N is large.

Thus convolution in general, including cross correlation and autocorrelation, is now best performed by taking the two DFT's, multiplying, and inverting the DFT. Some special considerations arise. Consider first a case where (Fig. 11.10) the two inputs f and g to be convolved have the same number of elements, as happens with autocorrelation. The output function will have twice as many elements as the input functions. So if one simply multiplies F and G term by term and retransforms, the output will be only half the correct length. The result of this procedure can be visualized in terms of cyclic convolution as in the example of autocorrelation shown in Fig. 11.4. What will happen is that the output sequence will close around the circle and overlap itself. Clearly this can be avoided, as in the figure, by packing the given functions with zeros; enough to double the length of the given sequences will suffice. Figure 11.11 brings out these practical points in a way that is glossed over in Fig. 11.9.

TIMING DIAGRAMS

Much literature devoted to fast algorithms has relied on numerical evaluation of computational complexity, usually in the form of counts of multiplications and additions, counts that are easily obtained by inserting a program line that increments a counter by one at each point in the program where the operation to be counted occurs. Such counts, as a function of N in the case of Fourier transforms, have frequently been published and compared with the results for predecessor programs. Programs that succeed in reducing the number of operation counts are

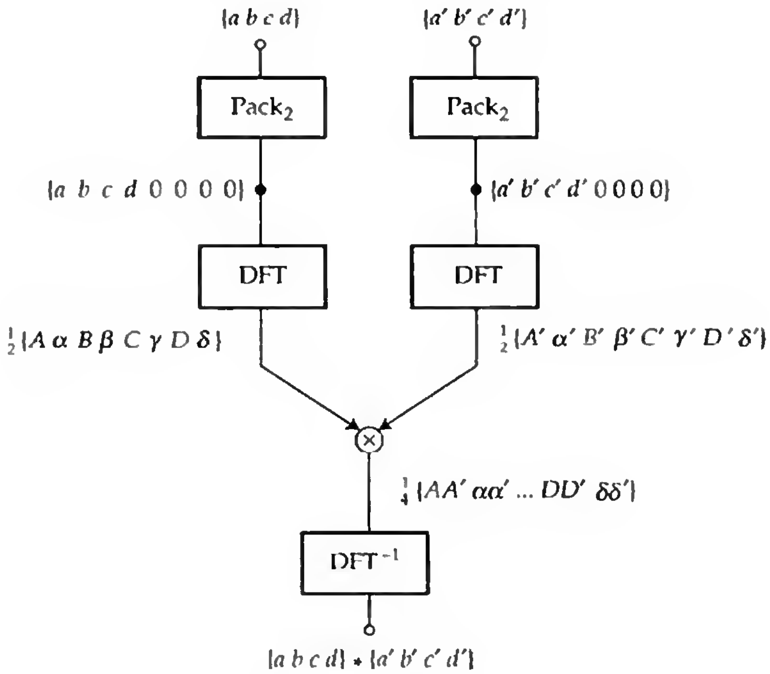


Fig. 11.11 **Convolution of two four-element sequences performed by using the DFT.**

then said to be faster. However, timing is the fundamental criterion for speed and has proved a convenient tool for investigating the time-consuming components such as permutation, pretabulation of sines and cosines, and other steps, as shown in Fig. 12.5.

MATLAB facilitates timing of program segments by insertion of **`tic`**, which starts a timer, and **`toc`**, which gives elapsed time. Use of this tool allows overall running time to be broken down into components to reveal steps where time is wasted. In the past, users of central computers found timing to be interfered with by time-sharing between users, and later even stand-alone computers found ways to perform clean-up tasks while ostensibly working on the owner's program. Such interference with timing, if present, can be revealed by rerunning the program and noting that the elapsed time is reported differently. Serious programmers can attempt to work around this unwanted behavior by plotting elapsed time against some parameter, such as N, where a smooth plot is expected, and superposing numbers of reruns to help identify spurious contributions to elapsed time.

WHEN N IS NOT A POWER OF 2

A data sequence of $N = 365$ elements is not directly amenable to successive halvings, but since $365 = 5 \times 73$ there is still, in theory, an advantage to be extracted. In practice, special base-3, base-5, base-7, base-11, and base-13 algorithms are avail-

able (see Nussbaumer, 1982, and Elliott and Rao, 1982, for algorithms due to S. Winograd), but no base-73 algorithm is to be expected. With 365 data values the straightforward procedure is to extend the sequence to 512 with appended zeros. Of course, the resulting 512 transform values will not refer to frequencies that are harmonics of one cycle per year. However, if needed, the 365 complex values will be delivered by the MATLAB `fft( )` operator.

TWO-DIMENSIONAL DATA

Let us compare the standard form of the two-dimensional Fourier transform (Chapter 13)

$$F(u,v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)e^{-i2\pi(ux+vy)}\,dx\,dy$$

with the two-dimensional discrete Fourier transform

$$F(\mu,\nu) = M^{-1}N^{-1}\sum_{\sigma=0}^{M-1}\sum_{\tau=0}^{N-1} f(\sigma,\tau)e^{-i2\pi(\mu\sigma/M+\nu\tau/N)}.$$

The integers σ and τ may be connected with the (x,y)-plane as follows. If the sampling intervals are X and Y, and $x_{\min}$ and $y_{\min}$ are the minimum values of x and y to be considered, then

$$\sigma = \frac{x - x_{\min}}{X}$$

$$\tau = \frac{y - y_{\min}}{Y}.$$

Since $M - 1$ and $N - 1$ are the largest values reached by σ and τ, respectively, it follows that

$$x_{\max} = x_{\min} + (M-1)X$$
$$y_{\max} = y_{\min} + (N-1)Y.$$

The spatial frequency integers μ and ν are such that μ/N and ν/M are spatial frequencies measured in cycles per sampling interval of x and y and μ/NX and ν/NY are spatial frequencies measured in cycles per unit of x and y. This discussion pictures $f(\sigma,\tau)$ as a function that possesses values in between its discrete samples but presumes those values to be unavailable. That situation often arises, which is why a connection with the integral transform has been established here. But we also understand that it is not necessary to regard $f(\sigma,\tau)$ as other than a function of integer pairs only and that μ and ν need not be regarded as frequencies. In fact, as previously noted in one dimension, great care has to be taken in interpreting μ and ν as frequencies.

Whereas the integral transform covers positive and negative areas of the (x,y)-plane, the discrete transform does not require negative values of σ and τ. Con-

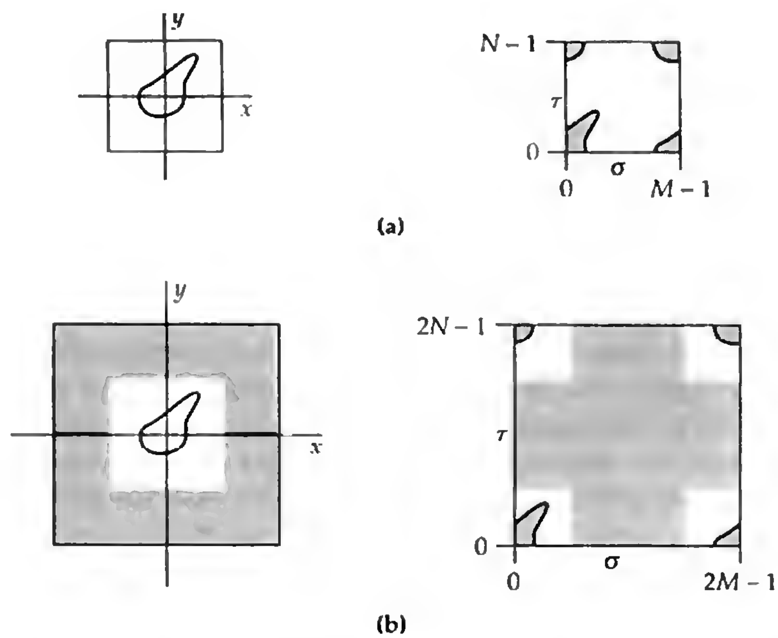


Fig. 11.12 **(a) When $f(x,y)$ is formed into an M × N array with positive subscripts starting from zero, it appears on the (σ,τ)-plane dissected as shown; (b) a surround of zeros in the (x,y)-plane (shown shaded) is not entered as a surround in the (σ,τ) plane.**

sequently, a simple object in the (x,y)-plane as in Fig. 11.12*a* becomes carved up in a strange way on the (σ,τ)-plane when the shifts of origin are made. It is very helpful to have the topology of this figure in mind when handling two-dimensional data. For example, the idea of surrounding the object with a guard zone of zeros, shown crosshatched, a trivial move in the (x,y)-plane, not requiring reassignment of existing data, calls for a rather tricky maneuver on the (σ,τ) matrix as shown in Fig. 11.12*b*.

If, by oversight or design, the σ and τ axes are not taken to coincide with the nominal axes x and y, then the transform obtained will be affected. For example, an object symmetrical with respect to the x and y axes will have a real-valued transform, but complex values will result if the σ and τ axes are shifted. If, not realizing this, one reads out the real transform values only, they will be wrong. However, if one reads out the complex values they will differ only in a trivial way from the nominal transform. The modulus will be correct and the phase will advance linearly with σ and τ in the way controlled by the shift theorem.

POWER SPECTRA

In many situations where transform phase is unimportant or unknowable, one could deal with $|F(\nu)|$, but it is customary to deal with $|F(\nu)|^2$, which is equiva-

lent, and to refer to $|F(\nu)|^2$ as the power spectrum. Values of $|F(\nu)|^2$ may indeed represent a number of watts in some applications, but even where the physical significance is not power or where there is no physical significance at all, the term "power spectrum" is in common use. The term is also used in connection with the Fourier transform $S(f)$ of a function of continuous time (pp. 120, 122), but there is a distinction. A power spectrum $|S(f)|^2$ would be measured in units of watts per hertz or something more complicated (such as ohm-watts per hertz as in Rayleigh's theorem on p. 206) but never in plain watts.

At first sight it might seem unnecessary to give special computational attention to the power spectrum, which is, after all, included within the larger concept of the (complex) Fourier transform. But, in fact, a great deal has been written under the heading of power spectra, where *power* spectra per se have not actually been of the essence. The literature referred to might equally well have been entitled spectra of random functions of time. The reason for the terminology is that random functions or noise waveforms present one of the important situations where phase loses meaning and the power spectrum becomes the natural entity to work with.

Although the power spectrum has a wider range of application than just to random processes or to deterministic signals of random origin, these applications are nevertheless important. The power spectrum of a random *process* is quite often defined as the Fourier transform of the autocovariance function of the process. (The autocovariance is what results when any d-c component is subtracted before autocorrelation is performed. But the distinction in terminology is not universally observed because it is commonly understood that any nonzero mean level is to be subtracted before calculating the autocorrelation, since the calculation is impossible otherwise.)

A definition of power spectrum in terms of autocorrelation (or autocovariance) seems indirect to many students but does permit the solution of problems involving random processes that are specified in the time domain by probabilities. When computation is required, however, one is never dealing with a random *process* but with an actual data string, possibly of random origin in some sense. (There may, in addition, be random errors of measurement.)

The curious fact that makes computation of power spectra so interesting is this. Suppose one takes the DFT of a data string of N elements. To be concrete, let the N data values be the height of the sea surface at a certain point, taken at 10-second intervals. Naturally, the values of $F_n(\nu)$ ought to show in what frequency bands the wave power resides, but the precision will be limited, as will the resolution, because N is only finite. The values of pha $F_N(\nu)$ will not be zero, but can hardly be expected to contain anything of interest, and if they are abandoned, $|F_N(\nu)|^2$ will constitute our measurement of the power spectrum of the waves. If the state of the sea were to change, as it is always doing, that measurement would have to stand as the record of the sea spectrum at that epoch. But because of the finite value of N, the measurements are imperfect to a degree that is evidenced by irregular variation from one value of ν to the next. To make a better measurement next time we might take data for four times as long, but how would we

know that the sea spectrum had not changed during the period of observation? The only way to tell from the data would be to divide the string into segments and make a judgment. Since this discussion has been cast in terms of sea waves, it is apparent that discussion of limits as $N \to \infty$ would be inappropriate, but so would it be in the case of almost all kinds of data. One can conceive of exceptions such as determination of the power spectrum of the data string constituted by the consecutive digits of the decimal expansion of π, but in the physical world things change if an observation takes too long. Yet experience might suggest that quadrupling N, thus staying far short of infinity, would double our precision or nearly so (i.e., the irregular variation might be halved). The strange thing is that the precision is not improved at all by increasing N. Thus, even in theory, the idea of defining the power spectrum as a limit as $N \to \infty$, does not work for data of random origin.

One might suspect that the paradox disappears if the situation is viewed in the right way. Here is an explanation. In any frequency band, chosen in advance, the amount of power will indeed be measured with increased precision as N is increased. If N is quadrupled the DFT will supply four values in a fixed frequency band that previously contained only one. Even though these four values are no more precise than the previous one, the sum of the four, which represents the new measurement of the power in the band, will have greater precision.

This correct view illuminates the procedure followed for computing power spectra of real data. The computed values of $|F_N(\nu)|^2$ will fall above and below some general trend with ν, and the scatter may be reduced by averaging several adjacent values. If high precision is sought by averaging too many consecutive values, the precision is paid for by loss of resolution in frequency, so a compromise must be arrived at by judgment based on experience with the character of the data. No unique advice can be offered by theory alone; that is why a variety of prescriptions can be found in the literature. In any case, the outcome is to smooth the sequence $|F(\nu)|^2$ by taking running means over a certain number of values, that is, by discrete convolution in the power spectrum domain with some sequence of weights.

Naturally several smoothing sequences have been proposed, but which is optimum? The answer to this depends on the purpose of the analysis and on the character of the data. Although it is true that smoothing increases precision, there are a number of accompanying effects that may be undesirable. For example, if there is a narrow spectral feature that is of interest, then extra smoothing will give an erroneous low value for the central strength, an erroneous large value for the width and may introduce lobe structure on each side. There are cases where absolute strength measurement is important as in chemical spectral analysis performed by Fourier transform spectroscopy, other cases where it is important to separate close spectral features, and others where there is a heavy penalty for false detection of faint features. In the latter case one may suppress lobes that might be counted as real and accept the accompanying loss of resolution as represented by widening of the spectral feature. In another case one might accept lobe structure in order to get a better strength measurement at a frequency peak. Even when

such costs and benefits are balanced to the user's satisfaction, the result will not necessarily be optimum for a new batch of data. It is apparent that selection of smoothing sequences goes beyond the realm of mathematical analysis to involve experience and judgment.

In principle, smoothing in the power spectrum domain is achievable by multiplying the autocorrelation function by a tapering factor. The term *lag window* for such a factor applied to the autocorrelation function was introduced by Blackmann and Tukey (1958). (*Spectral window* is the Fourier transform of the lag window.) However, when large amounts of data are involved, it is convenient to compute the autocorrelation by first performing the FFT, then forming the power spectrum and inverting the FFT. Once the power spectrum is formed, one might as well do the smoothing directly, especially as the sequence of weights is only likely to be short.

In an earlier section it was shown that the DFT of a sampled function is not necessarily in close agreement with the Fourier transform of the function itself, especially where discontinuities in the function cause the spectrum to run out to high frequencies. A finite segment of data from a data stream will possess exactly the sharp start and finish that give rise to such disagreement. Therefore, some practitioners apply a tapering factor to the data to remove or reduce the terminal discontinuities. This is not the same thing as a lag window, but the general effect will be a "power spectrum" that is smooth. In general, the result will not be expressible as a convolution between the raw power spectrum and a sequence of weights and therefore does not strictly represent the power in a frequency band. Whether this cost is less than the cost of providing for zeros must depend on the case.

Clearly there is more to using the FFT than a prepackaged software program can provide. Because circumstances are varied, and users differ according to their needs, use of the FFT is rather an art. As programs have become very fast and convenient, experimenting with one's own data and desiderata in order to gain relevant experience has become a valuable use for commercial programs.

BIBLIOGRAPHY

Bergland, G. D.: "A Guided Tour of the Fast Fourier Transform," *IEEE Spectrum*," vol. 6, pp. 41–52, 1969.

Blackman, R. B., and J. W. Tukey: "The Measurement of Power Spectra," Dover Publications, New York, 1958.

Bracewell, R. N.: "The Hartley Transform," Oxford University Press, New York, 1986.

Bracewell, R. N.: "Two-Dimensional Imaging," Prentice-Hall, Englewood Cliffs, New Jersey, 1995.

Brigham, E. O.: "The Fast Fourier Transform," Prentice-Hall, Englewood Cliffs, New Jersey, 1988.

Cochran, W. T., et al.: "What Is the Fast Fourier Transform?" *IEEE Trans. Audio Electroacoustics*, vol. AU-15, pp. 45–55, 1967.

Cooley, T. W., and J. W. Tukey: "An Algorithm for the Machine Calculation of Complex Fourier Series," *Math. Computation*, vol. 19, pp. 297–301.

Elliott, D. F., and K. R. Rao: "Fast Transforms: Algorithms, Analyses, Applications," Academic Press, New York, 1982.

Ersoy, O. K.: "Fourier-related Transforms," Prentice-Hall, Upper Saddle River, New Jersey, 1997.

Evans, D. M.: "An Improved Bit-reversal Algorithm," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. ASSP-35, pp. 1120–1125, 1987.

Gertner, I.: "A New Efficient Algorithm to Compute the Two-Dimensional Fourier Transform," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. ASSP-36, pp. 1036–1050, 1988.

Good, I. J.: "Random Motion on a Finite Abelian Group," *Proc. Camb. Phil. Soc.*, vol. 47, pp. 756–762, 1951.

Nussbaumer, H. J.: "Fast Fourier and Convolution Algorithms," Springer-Verlag, Berlin, 1982.

Oppenheim, A. V., and R. W. Schafer: "Discrete-time Signal Processing," Prentice-Hall, Englewood Cliffs, New Jersey, 1989.

Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling: "Numerical Recipes, the Art of Scientific Programming," Cambridge University Press, Cambridge, UK, 1990. See also successor volumes devoted to programs and examples in BASIC, C, FORTRAN 77, FORTRAN 90 and PASCAL.

Rabiner, L. R., and B. Gold: "Theory and Application of Digital Signal Processing," Prentice-Hall, Englewood Cliffs, New Jersey, 1975.

Rao, K. R., and P. Yip: "Discrete Cosine Transform," Academic Press, New York, 1990.

———. Special issue on the fast Fourier transform with an extensive bibliography. *IEEE Trans. Audio and Electroacoustics*, vol. AU-17, no. 2, June 1969.

Walker, J. S.: "Fourier Analysis," Oxford University Press, New York, 1988.

Walker, J. S.: "A New Bit-reversal Algorithm," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. ASSP-38, pp. 1472–1474, 1990.

PROBLEMS

1. **Discrete representation.** A function $(1 + t^2)^{-1}\Pi(t/10)$ is to be represented discretely for the purpose of computer experiments on the discrete Fourier transform using $N = 16$ samples. Draw up a suitable table of values.

τ	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$f(\tau)$																

2. **Cyclic convolution.** Obtain the following cyclic convolution sums:

$$\{1\,2\,0\,0\} * \{2\,3\,0\,0\},\ \{1\,1\,1\,1\} * \{0\,1\,1\,0\},\ \text{and}\ \{1\,0\,0\,1\} * \{0\,1\,0\,0\}.$$

3. **Discrete transforms.** Obtain the DFT of the following sequences, checking the results by both the "sum of sequence" and "first-value" rules:

$$\{1\,2\,3\,4\} \quad \text{and} \quad \{1\,2\,3\,4\,0\,0\,0\,0\}.$$

4. **Convolution theorem.** Obtain the cyclic convolutions

$$\{0\,1\,0\,0\} * \{0\,0\,1\,0\} \quad \text{and} \quad \{1\,1\,0\,0\} * \{0\,0\,1\,1\}$$

and verify the results by the convolution theorem.

5. **Two-dimensional convolution.** Obtain the two-dimensional cyclic convolution sums

$$\begin{Bmatrix} 0 & 0 \\ 1 & 0 \end{Bmatrix} ** \begin{Bmatrix} 0 & 0 \\ 0 & 1 \end{Bmatrix} \quad \text{and} \quad \begin{Bmatrix} 1 & 0 \\ 0 & 2 \end{Bmatrix} ** \begin{Bmatrix} 0 & 1 \\ 0 & 0 \end{Bmatrix}$$

and verify that the results are correct by mean of the two-dimensional convolution theorem.

6. **Two-dimensional DFT.** Verify the following four DFT pairs in two dimensions and show that the fifth results from adding the four together.

$$\begin{Bmatrix} 0 & 0 \\ 1 & 0 \end{Bmatrix} \supset \frac{1}{4}\begin{Bmatrix} 1 & 1 \\ 1 & 1 \end{Bmatrix}$$

$$\begin{Bmatrix} 0 & 0 \\ 0 & 1 \end{Bmatrix} \supset \frac{1}{4}\begin{Bmatrix} 1 & -1 \\ 1 & -1 \end{Bmatrix} \qquad \begin{array}{l} \tau \\ \uparrow \\ \{\quad\} \rightarrow \sigma \end{array}$$

$$\begin{Bmatrix} 1 & 0 \\ 0 & 0 \end{Bmatrix} \supset \frac{1}{4}\begin{Bmatrix} -1 & -1 \\ 1 & 1 \end{Bmatrix}$$

$$\begin{Bmatrix} 0 & 1 \\ 0 & 0 \end{Bmatrix} \supset \frac{1}{4}\begin{Bmatrix} -1 & 1 \\ 1 & -1 \end{Bmatrix} \qquad \begin{array}{l} \nu \\ \uparrow \\ \{\quad\} \rightarrow \mu \end{array}$$

$$\begin{Bmatrix} 1 & 1 \\ 1 & 1 \end{Bmatrix} \supset \frac{1}{4}\begin{Bmatrix} 0 & 0 \\ 4 & 0 \end{Bmatrix}$$

Enunciate two-dimensional sum and first-value theorems. Reconcile the pattern of minuses in the transform with what might be expected from the two-dimensional shift theorem. Verify the two-dimensional Rayleigh-Parseval theorem on the examples.

7. **Midpoint interpolation.** A 16-element sequence whose DFT is $F(\nu)$ is extended to 32 elements by the addition of 16 trailing zeros. The new DFT is $G(\nu)$. By the packing theorem we can obtain half the values of $G(\nu)$ immediately; for example, $G(0) = 0.5F(0)$, $G(2) = 0.5F(1)$, $\ldots$ $G(30) = 0.5F(15)$. Show that the intermediate values of $G(\nu)$ can be obtained by cyclic convolution of the known values with the midpoint interpolation sequence (p. 225), that is, that

$$G(\nu) = 0.5 \sum_{\kappa=-\infty}^{\infty} F\left(\frac{\nu}{2} - \kappa - \tfrac{1}{2}\right) \operatorname{sinc}\left(\kappa + \tfrac{1}{2}\right), \qquad \nu = 1, 3, \ldots, 31.$$

8. **Supplement to packing theorem.** The preceding problem requires summing an infinite number of terms to obtain intermediate values of $G(\nu)$. Show that the sum reduces to a finite number of terms if the interpolating coefficients are modified as follows:

$$G(\nu) = 0.5 \sum_{\kappa=0}^{N-1} F\left(\frac{\nu}{2} - \kappa - \tfrac{1}{2}\right) \frac{\sin\left[\pi\left(\kappa + \frac{1}{2}\right)\right]}{\sin\left[\pi N^{-1}\left(\kappa + \frac{1}{2}\right)\right]}, \qquad \nu = 1, 3, \ldots, 2N - 1,$$

where N is the number of elements in the original sequence before the N trailing zeros are added.

9. **Two-dimensional autocorrelation.**
   (*a*) Calculate the two-dimensional autocorrelation of the following data array (non-cyclically):

$$\begin{array}{cc|cc} & 1 & 1 & 1 \\ \tau & & & \\ & 0 & 1 & 2 \\ \hline & & 0 & 1 \\ & & \sigma & \end{array}$$

   (*b*) Describe the nature of the symmetry possessed by the result.
   (*c*) Calculate the autocorrelation of the following one-dimensional sequences $\{1\ 1\ 1\ 2\}$, $\{1\ 1\ 0\ 1\ 2\}$, and $\{1\ 1\ 0\ 0\ 1\ 2\}$. Can you see a way of performing the two-dimensional autocorrelation by doing only one-dimensional autocorrelation?

10. **Comparison of DFT with FT.** A function of the discrete variable τ is defined by $f(\tau) = \exp(-\tau/4)$ for $1 \leq \tau \leq 31$, and for $\tau = 0$ the initial value $f(0) = 0.5$. Compute the DFT and compare with $V(f)$ the FT of $v(t) = \exp(-t/4)H(t)$. ▷

11. **Smoothing due to cosine-bell taper.** If a function $g(x)$ is gated by multiplication with $\Pi(x/X)$ and then tapered by multiplication by a cosine bell $0.5 + 0.5\cos(2\pi x/X)$ its Fourier transform $G(s)$ is smoothed as by convolution with $X \operatorname{sinc} Xs$ $[\frac{1}{4}\delta(s + X^{-1}) + \frac{1}{2}\delta(s) + \frac{1}{4}\delta(s - X^{-1})]$. If an N-element sequence $f(\tau)$ is similarly tapered so as to rise from zero at the beginning ($\tau = 0$) and fall to zero again at the end ($\tau = N$), show that the discrete transform $F(\nu)$ will be modified by discrete convolution with the N-element sequence $\{[\frac{1}{2}\ -\frac{1}{4}\ 0 \ldots 0\ -\frac{1}{4}]\}$. The effect of this operation is to replace any value $F(\nu)$ by $-\frac{1}{4}F(\nu - 1) + \frac{1}{2}F(\nu) - \frac{1}{4}F(\nu + 1)$. From Problem 3.1*f* (p. 58) we know that serial products with $\{\frac{1}{4}\ \frac{1}{2}\ \frac{1}{4}\}$ lead to smoothing. Would not the sequence $\{-\frac{1}{4}\ \frac{1}{2}\ -\frac{1}{4}\}$ lead to sharpening?

12. **Index reversal in MATLAB.** A row vector $\{1\,2\,3\,4\,5\,6\,7\}$ is to be reversed to form a new row vector $y(n) = \{1\,7\,6\,5\,4\,2\,2\}$. Show how to perform the reversal operation in MATLAB.

13. **M-file for the discrete Hartley transform.** (*a*) We would like to have a MATLAB function named **dht** that will operate on an 8-element data sequence such as $\{1\,2\,3\,4\,5\,6\,7\,8\}$, for instance, and produce the known discrete Hartley transform.

$$\{4.5000 \quad -1.7071 \quad -1.0000 \quad -1.7071 \quad -0.5000 \quad -0.2929 \quad 0.0000\}.$$

The desired function program would read as follows

```
f=[1 2 3 4 5 6 7 8];
N= 8;
dht(f,N)
```

Write the M-file whose first line reads **`function y = dht(f,N).`**

(*b*) With the command **`dht`** available, compare the discrete Hartley transform results with some of the discrete Fourier transforms given in Chapter 11. Execute **`dht(dht(f,N))`** for inspection and then write an M-file **`idht`** that delivers the inverse discrete Hartley transform. ▷

14. Discrete FT versus expectation. (*a*) Verify that the sequence {1 2 3 4 5 4 3 2 1} consists of samples at unit spacing of the continuous triangle function $5\Lambda(x/5)$ whose Fourier transform is $25\,\mathrm{sinc}^2\,5s$. (*b*) Obtain the DFTs of

{1 2 3 4 5 4 3 2 1},

{0 0 0 0 1 2 3 4 5 4 3 2 1 0 0 0 0},

{5 4 3 2 1 0 0 0 0 0 0 0 0 0 0 0 0 1 2 3 4},

using the DFT definition sum or an available built-in routine such as **`fft(  )`** in MATLAB. Which, if any, of the DFTs obtained meets the expectation of agreement with $25\,\mathrm{sinc}^2\,5s$? ▷

15. DFT of binomial coefficients. Take {1 4 6 4 1} as an example of the binomial coefficients $B(n,x) = n!/(n-x)!x!$ for $n = 4$ and fit a normal distribution $A\exp(-x^2/2\sigma^2)$. Match σ to the root-mean-square departure of x from its mean (namely $\sigma = 1$), and fix A by equating the area under the normal curve to the sum of the coefficients. Obtain the DFT of {6 4 1 0 0 0 0 0 0 0 0 0 0 0 1 4} and compare with the FT of $A\exp(-x^2/2\sigma^2)$. ▷

16. Cyclic convolution sum. Given two sequences {1 7 21 35 35 21 7 1} and {1 5 9 5 −5 −9 −5 −1}, what is their cyclic convolution on a 12-element circular support? ▷

17. A timing experiment. Use the definition of T4, the nominal time for four multiplies, to see whether the time changes when an average is taken over many more or many fewer than $N = 1000$ tries. If possible, try two different computers, or different languages. Find out what can be deduced from such timing experiments.

18. Sinc interpolation. Midpoint interpolation by convolution with a sequence of samples of the sinc function was described in Chapter 10. The same result is obtained by truncating $\{f\}$ to N elements and taking the DFT of $\{f\}$ to obtain $\{F\}$, which has N complex values. Now create a sequence $\{G\}$ that retains these N values but has an additional N trailing zeros to extend the length to $2N$ elements, the zeros being used to extend the transform to higher frequencies. Take the inverse transform, which will now have $2N$ values. Half of these agree with $\{f\}$, the other half being midpoint interpolates. (*a*) Write a short program to illustrate the method. (*b*) What adjustment is necessary to compensate for the change in N? (*c*) What should be done to interpolate three values between the given values of $\{f\}$? (*d*) Compare with other methods of sinc-function interpolation. ▷

12

The Discrete Hartley Transform[1]

A STRICTLY RECIPROCAL REAL TRANSFORM

Given a real waveform $V(t)$, we can define the integral transform

$$\psi(\omega) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{\infty} V(t)(\cos \omega t + \sin \omega t)\, dt, \tag{1}$$

where the integral exists. The waveform may be complex, but will be taken as real in what follows, and may contain generalized functions such as delta functions and their derivatives. Clearly $\psi(\omega)$ is a sum of double-sided sine and cosine transforms from whose reciprocal properties one readily deduces the inverse relation

$$V(t) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{\infty} \psi(\omega)(\cos \omega t + \sin \omega t)\, d\omega. \tag{2}$$

These relations (p. 200) were presented by Hartley (1942) and never disappeared from the technical literature, but have not been well known. The direct and inverse relations are identical in form and if the given waveform $V(t)$ is real, so is its waveform $\psi(\omega)$.

To connect the transform $\psi(\omega)$ with the Fourier transform $S(\omega)$ of $V(t)$, it pays us to adopt the definition

$$S(\omega) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{\infty} V(t) e^{-i\omega t}\, dt$$

[1] Reprinted with permission from *J. Opt. Soc. Am.*, vol. 73, no. 12, December 1983, pp. 1832–1835.

and its inverse

$$V(t) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{\infty} S(\omega) e^{i\omega t}\, d\omega.$$

Let $\psi(\omega) = e(\omega) + o(\omega)$, where $e(\omega)$ and $o(\omega)$ are the even and odd parts of $\psi(\omega)$, respectively. Then

$$e(\omega) = \frac{\psi(\omega) + \psi(-\omega)}{2} = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{\infty} V(t) \cos \omega t\, dt$$

and

$$o(\omega) = \frac{\psi(\omega) - \psi(-\omega)}{2} = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{\infty} V(t) \sin \omega t\, dt.$$

Given $\psi(\omega)$ we may form the sum $e(\omega) - io(\omega)$ to obtain the Fourier transform $S(\omega)$:

$$S(\omega) = e(\omega) - io(\omega) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{\infty} V(t)(\cos \omega t - i \sin \omega t)\, dt. \tag{3}$$

Thus we see that from $\psi(\omega)$ one readily extracts the Fourier transform of $V(t)$ by simple reflections and additions.

Conversely, given the Fourier transform $S(\omega)$, we may obtain $\psi(\omega)$ by noting that

$$\psi(\omega) = \mathrm{Re}\,[S(\omega)] - \mathrm{Im}\,[S(\omega)]. \tag{4}$$

Thus from $S(\omega)$ one finds $\psi(\omega)$ as the sum of the real part and sign-reversed imaginary part of the Fourier transform.

NOTATION AND EXAMPLE

As an example take

$$V(t) = \begin{cases} \exp(-t/2) & t > 0 \\ 0 & t < 0. \end{cases}$$

Then

$$S(\omega) = \frac{1 - i\omega}{1 + \omega^2}$$

and

$$\psi(\omega) = \frac{(2\pi)^{\frac{1}{2}}(1 + \omega)}{1 + \omega^2}.$$

Figure 12.1 shows $V(t)$ on the left and the Fourier transform $S(\omega)$ on the right; the real part of the transform is a broken line and the imaginary part is dotted. The imaginary part has been reversed in sign. Hartley's transform, shown by the full line, is simply the sum of the real part and the sign-reversed imaginary part

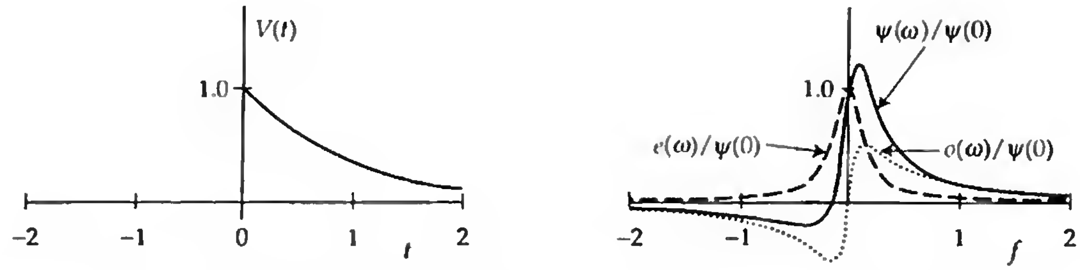


Fig. 12.1 **On the left a waveform $V(t)$ and on the right the real part of the spectrum $S(\omega)$ (broken), the sign-reserved imaginary part of $S(\omega)$ (dotted) and Hartley's transform of $V(t)$ (full line).**

of $S(\omega)$. It is real and clearly unsymmetrical. From its even and odd parts we could readily reverse the construction to recover the real and imaginary parts of the complex-valued Fourier transform $S(\omega)$.

For historical continuity we have retained the elegant factors $(2\pi)^{\frac{1}{2}}$ used by Hartley. But in what follows we drop these factors and move to the more familiar notation where the 2π appears only in the combination $2\pi \times$ frequency.

THE DISCRETE HARTLEY TRANSFORM

Now consider a discrete variable τ that is like time but can assume only the N integral values from 0 to $N - 1$. Given a function $f(\tau)$, which one could think of as the representation of a waveform, we define its DHT to be

$$H(\nu) = N^{-1} \sum_{\tau=0}^{N-1} f(\tau) \operatorname{cas}\left(\frac{2\pi\nu\tau}{N}\right), \tag{5}$$

where $\operatorname{cas} \theta = \cos \theta + \sin \theta$, an abbreviation adopted from Hartley. For comparison, the discrete Fourier transform $F(\nu)$ is

$$F(\nu) = N^{-1} \sum_{\tau=0}^{N-1} f(\tau) \exp\left(-\frac{i2\pi\nu\tau}{N}\right).$$

The inverse DHT relation is

$$f(\tau) = \sum_{\nu=0}^{N-1} H(\nu) \operatorname{cas}\left(\frac{2\pi\nu\tau}{N}\right). \tag{6}$$

To derive this result we use the orthogonality relation

$$\sum_{\nu=0}^{N-1} \operatorname{cas}\left(\frac{2\pi\nu\tau}{N}\right) \operatorname{cas}\left(\frac{2\pi\nu\tau'}{N}\right) = \begin{cases} N & \tau = \tau' \\ 0 & \tau \neq \tau'. \end{cases}$$

Substituting (5) into the right-hand side of (6),

$$\begin{aligned}\sum_{\nu=0}^{N-1} H(\nu)\operatorname{cas}\left(\frac{2\pi\nu\tau}{N}\right) &= \sum_{\nu=0}^{N-1} N^{-1}\sum_{\tau'=0}^{N-1} f(\tau')\operatorname{cas}\left(\frac{2\pi\nu\tau'}{N}\right)\operatorname{cas}\left(\frac{2\pi\nu\tau}{N}\right)\\ &= N^{-1}\sum_{\tau'=0}^{N-1} f(\tau')\sum_{\nu=0}^{N-1}\operatorname{cas}\left(\frac{2\pi\nu\tau'}{N}\right)\operatorname{cas}\left(\frac{2\pi\nu\tau}{N}\right)\\ &= N^{-1}\sum_{\tau'=0}^{N-1} f(\tau')\times\begin{cases}N & \tau=\tau'\\ 0 & \tau\neq\tau'\end{cases}\\ &= f(\tau),\end{aligned}$$

which verifies (6).

We see that the DHT is symmetrical, apart from the factor N^{-1} which is familiar from the DFT, and it is real.

The DHT could be made strictly symmetrical by changing the factor N^{-1} in the definition of the DHT to $N^{-1/2}$. Then changing the definition of the inverse DHT to include a factor $N^{-1/2}$ we would have a strictly symmetrical transformation. Hartley ensured the symmetry of his original integral relations by such means. However, it is often convenient when computing to stop short of multiplication throughout by the constant factor and simply allow it to be absorbed into a normalization factor or graphics scale factor that will be introduced later. To indicate this relatively benign asymmetry we may say that the DHT is quasi-symmetrical.

To get the DFT from the DHT, split the latter into its "even" and "odd" parts.

$$H(\nu) = E(\nu) + O(\nu),$$

where
$$E(\nu) = \frac{H(\nu) + H(N-\nu)}{2}$$

and
$$O(\nu) = \frac{H(\nu) - H(N-\nu)}{2}.$$

Then the DFT is given by

$$F(\nu) = E(\nu) - iO(\nu).$$

Conversely, $H(\nu) = \mathcal{F}[f_{\text{even}}] - \mathcal{F}[f_{\text{odd}}]$.

As the real, strictly invertible, integral transform which forms the basis for the present discrete formulation was originally presented in 1942 by Ralph V. L. Hartley, it is appropriate to name the DHT, derived from his idea, in his honor. Hartley was a Rhodes scholar, a Fellow of the Institute of Radio Engineers, and rose to be in charge of telephone line research at the Bell Telephone Laboratories (1918–1929). To generations of electrical engineers his name was well known through the Hartley oscillator, which was once the standard textbook example of an electronic sine-wave source. A simple system to analyze, it comprised a single triode amplifier with feedback derived from a tapped inductance. For a bibliog-

raphy of hundreds of papers see the special issue of *Proc. IEEE*, March 1994, with a special section devoted to Hartley transform (edited by K. J. Olejniczak and G. T. Herdt). For subsequent papers use the *Science Citation Index*.

EXAMPLES OF DHT

For comparison with Fig. 12.1, consider

$$f(\tau) = \begin{cases} 0.5 & \tau = 0 \\ \exp(-\tau/2) & \tau = 1, 2, \ldots, 15, \end{cases}$$

which represents the earlier function of continuous t by $N = 16$ equispaced samples. The value at $\tau = 0$, since it falls on the discontinuity of $V(t)$, is assigned as $[V(0+) + V(0-)]/2 = 0.5$. The result for $H(\nu)$, which is shown in Fig. 12.2, closely resembles samples of the transform of Fig. 12.1 taken at intervals $\Delta\omega/2\pi = 1/16$.

The discrepancies, which are small in this example, are due partly to the truncation of the exponential waveform and partly to aliasing, exactly as with the DFT. As a final example, take the binomial sequence 1, 6, 15, 20, 15, 6, 1 representing samples of a smooth pulse. To obtain the simplest result assign the peak value at $\tau = 0$. Thus

τ, ν	=	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15,
$f(\tau)$	=	20	15	6	1	0	0	0	0	0	0	0	0	0	1	6	15,
$H(\nu)$	=	4	3.56	2.49	1.32	0	0.12	0.01	0	0	0	0	0.12	0.5	1.32	2.49	3.56.

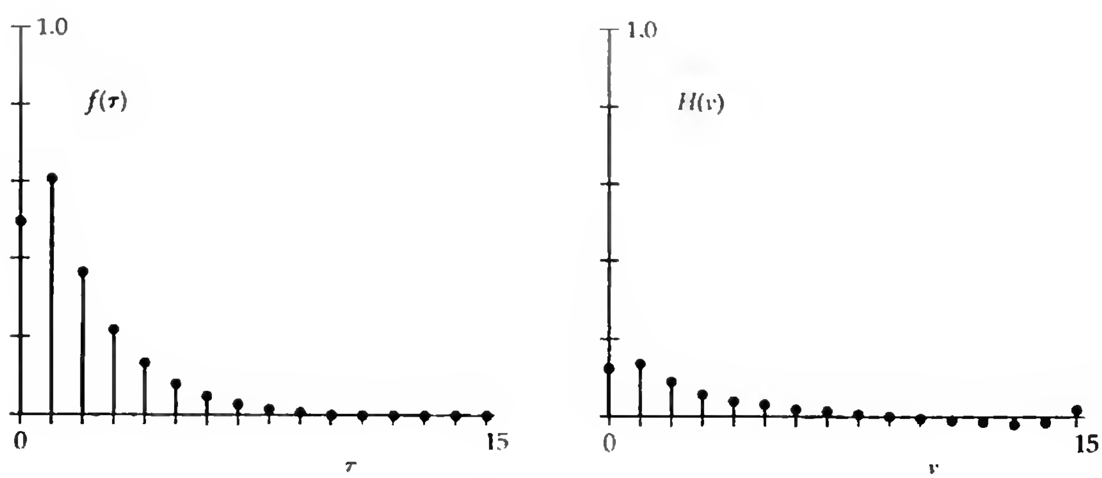


Fig. 12.2 A 16-point representation of the truncated exponential waveform as used in Fig. 12.1 (left) and its DHT (right).

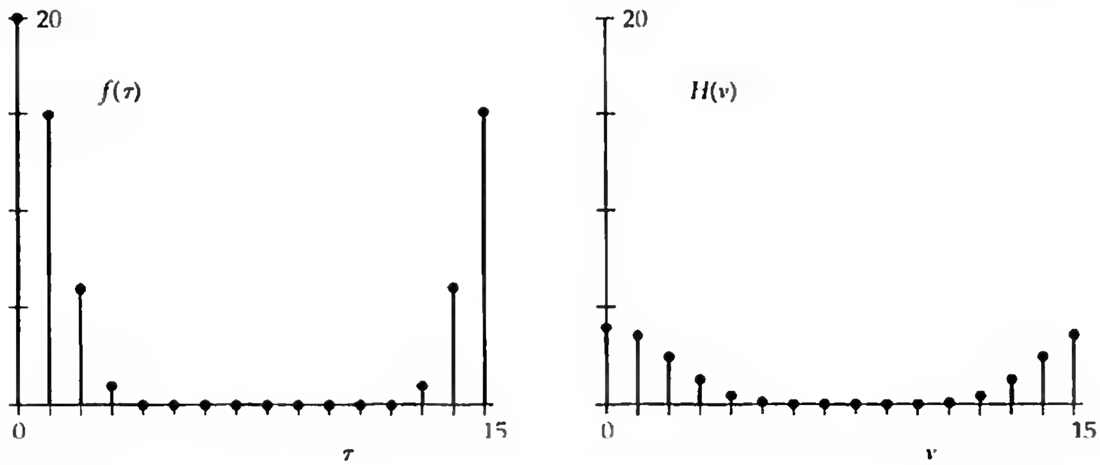


Fig. 12.3 **Samples representing a smooth binomial hump (left) and its DHT (right).**

The result shown in Fig. 12.3 is the expected smooth pulse peaking at $\nu = 0$.

For numerical checking it is useful to know that, as for the discrete Fourier transform, the sum of the DHT values $\Sigma H(\nu)$ is equal to $f(0)$. Conversely, the sum of the data values $\Sigma f(\tau)$ is equal to $NF(0)$.

DISCUSSION

At first sight it may seem strange that N real values of the DHT can substitute for the N complex values of the DFT, a total of $2N$ real numbers. We can understand this, however, by remembering that the Hermitian property of the DFT means redundancy by a factor of 2. The $N/2$ real numbers that suffice to specify the cosine transform combine with the $N/2$ needed for the sine transform to form a total of N DHT coefficients containing no degeneracy due to symmetry.

The function cas θ, which may be thought of as a sine wave shifted 45°, automatically responds to cosine and sine components equally. If, as a kernel, we use $2^{\frac{1}{2}} \sin(\theta + \alpha)$, where α is an arbitrary shift, the responses will be unequal but no information will be lost unless $\alpha = 0, \pi/2, \ldots$. Consequently one would expect to be able to invert; the inversion kernel is $\cot^{\frac{1}{2}} \alpha \sin \theta + \tan^{\frac{1}{2}} \alpha \cos \theta$.

A CONVOLUTION ALGORITHM IN ONE AND TWO DIMENSIONS

The convolution theorem obeyed by the DHT is as follows. If $f(\tau)$ is the convolution of $f_1(\tau)$ with $f_2(\tau)$, i.e.,

$$f(\tau) = f_1(\tau) * f_2(\tau) = \sum_{\tau'=0}^{N-1} f_1(\tau') f_2(\tau - \tau'),$$

then $$H(\nu) = H_1(\nu)H_{2e}(\nu) + H_1(-\nu)H_{2o}(\nu),$$

where $H(\nu)$, $H_1(\nu)$, and $H_2(\nu)$ are the DHTs of $f(\tau)$, $f_1(\tau)$, and $f_2(\tau)$, respectively, and $H_2(\nu) = H_{2e}(\nu) + H_{2o}(\nu)$, the sum of its even and odd parts.

A customary way of performing convolution numerically is to take the discrete Fourier transform of each of the two given sequences and then to do complex multiplication element by element in the Fourier domain, which will require four real multiplications per element. Then one inverts the Fourier transform, remembering to change the sign of i. Analogous use of the Hartley transform for convolution will require only two real multiplications per element.

However, it very often happens, especially in image processing but also in digital filtering in general, that one of the convolving functions, say $f_2(\tau)$, is even. In that case $H_{2o}(\nu)$ is zero, and the convolution theorem therefore simplifies to $H(\nu) = H_1(\nu)H_2(\nu)$. In view of this simplified form of the theorem one need only take the DHTs of the two data sequences, multiply together term by term the two real sequences resulting, and take one more DHT. Thus the proposed procedure is

$$f_1(\tau) * f_2(\tau) = \text{DHT of } [(\text{DHT of } f_1) \times (\text{DHT of } f_2)].$$

Before taking the DHT, one extends the given sequences to twice the original length with zeros in order to have space for the convolution. The zeros may lead, trail, or bracket the data. When the sequences to be convolved are of unequal but comparable length, the method of extending with zeros and multiplying transforms may be used; but if one sequence is much shorter than the other, as is usual in digital filtering, then direct convolution is recommended (see Chapter 3 for simple code).

For image processing in two dimensions exactly the same procedure is available and the advantages of avoiding complex arithmetic and nonreciprocal subprograms are increased.

TWO DIMENSIONS

Manipulation of two-dimensional images may also benefit from the existence of a real transform. An image $f(\tau_1,\tau_2)$ represented by an $M \times N$ matrix does indeed possess a two-dimensional discrete Hartley transform (^{2}DHT) which is itself an $M \times N$ matrix $H(\nu_1,\nu_2)$ of real numbers. The transformation and its inverse are

$$H(\nu_1,\nu_2) = M^{-1}N^{-1} \sum_{\tau_1=0}^{M} \sum_{\tau_2=0}^{N} f(\tau_1,\tau_2) \operatorname{cas}[2\pi(\nu_1\tau_1 + \nu_2\tau_2)],$$

$$f(\tau_1,\tau_2) = \sum_{\nu_1=0}^{M} \sum_{\nu_2=0}^{N} H(\nu_1,\nu_2) \operatorname{cas}[2\pi(\nu_1\tau_1 + \nu_2\tau_2)].$$

The two-dimensional Hartley transform can be broken down into N one-dimensional transforms as explained by Bracewell et al. (1986) or can be done more efficiently (Meher, 1992; Yang, 1989).

A convenience of the Hartley transform in two dimensions is that it may be presented as a single diagram representing the full spectral analysis of an image, whereas the Fourier transform of an image, which is hardly ever published, requires two diagrams, either for the real and imaginary parts, or for the amplitude and phase. Use of this presentation in optics is exemplified in Bracewell and Villasenor (1990). A striking example of the Hartley transform of a spiral slit, prepared by J. Villasenor, appears on the cover of the *Proceedings of the Institute of Electrical and Electronic Engineers*, March 1994, while on pp. 587 and 588 of Bracewell (1995b) a complex Fourier transform of a 2D object can be contrasted with the single real Hartley diagram that contains the full phase and amplitude information about the object. Computer storage of a two-dimensional transform as a single real array has the same convenience as the graphical representation. Where phase is not included in the presentation, the two-dimensional power spectrum is the same for both transforms.

For three dimensions, see Hao and Bracewell (1987), and for four dimensions see Buneman (1987).

THE Cas-Cas TRANSFORM

In two dimensions, instead of the kernel $\text{cas}[2\pi(\nu_1\tau_1 + \nu_2\tau_2)]$ one may use $\text{cas}\, 2\pi\nu_1\tau_1 \,\text{cas}\, 2\pi\nu_2\tau_2$, the distinguishing feature of which is to be separable (Perkins, 1987). The cas-cas transform is reciprocal (Bracewell, 1983) and is related to the standard two-dimensional Hartley transform in the same way that arises when choosing between Fourier components of the form $\cos 2\pi[(\nu_1\tau_1 + \nu_2\tau_2)]$, representing corrugations on the (ν_1,ν_2)-plane with parallel null loci, and those of the form $\cos 2\pi\nu_1\tau_1 \cos 2\pi\nu_2\tau_2$, whose null loci intersect at right angles. The choice of kernel has also been discussed by Millane (1994) who gives additional references. Separability has not proved to be a basis for preference; consequently the kernel $\text{cas}[2\pi(\nu_1\tau_1 + \nu_2\tau_2)]$ has become the standard.

THEOREMS

There is a Hartley transform theorem for every theorem that applies to the Fourier transform, some of the theorems corresponding exactly, as with $\Sigma H(\nu) = f(0)$ and $\Sigma f(\tau) = NH(0)$. Likewise, the Hartley transform of most convolutions is the product of the separate Hartley transforms, as mentioned above. In other cases there are differences. For example, the shift theorem, needed in implementation of the fast Hartley algorithm, is

$$\text{DHT of } f(\tau + a) = H(\nu)\cos\left(\frac{2\pi a\nu}{N}\right) - H(-\nu)\sin\left(\frac{2\pi a\nu}{N}\right).$$

There is a quadratic content theorem

$$\sum_{\tau=0}^{N-1}[f(\tau)]^2 = N\sum_{\nu=0}^{N-1}[H(\nu)]^2$$

which resembles the analogous theorem for the discrete Fourier transform except that no complex conjugates enter, only real numbers.

According to the reversal theorem, $f(-\tau)$ has DHT $H(-\nu)$, where the negative arguments are interpreted modulo N; that is, where the argument falls outside the range 0 to $N-1$, add or subtract multiples of N as needed.

The first difference theorem states that $f(\tau+1)-f(\tau)$ has DHT $[\cos(2\pi\nu/N)-1]H(\nu)-\sin(2\pi\nu/N)H(N-\nu)$.

THE DISCRETE Sine AND Cosine TRANSFORMS

Many examples of the integral cosine transform of $f(x)$, defined earlier as $2\int_0^\infty f(x)\cos 2\pi sx\,dx$, are obtainable directly from the Pictorial Dictionary of Fourier Transforms by looking for purely even graphs, in either column. In addition, cosine transforms can be arrived at algebraically by extending a function, given for $x \geq 0$, to the left so as to generate an even function and then applying any of the various techniques for obtaining Fourier transforms. Were it not for extensive tables (notably Erdélyi's Tables of Integral Transforms) that exist for Fourier sine and cosine transforms of functions from advanced mathematics, these transforms might be regarded as merely special cases of the Fourier transform.

The cosine transform has a discrete version $\sum_0^{N-1} f(\tau)\cos(2\pi\nu\tau/N)$ (and likewise for the sine transform), but this discrete transform cannot be inverted to recover $f(\tau)$ because components of $f(\tau)$ of the form $\sin(\pi\nu\tau/N)$ are integrated out and leave no trace. The sine and cosine transforms themselves are invertible, but only when attention is restricted to functions $f(x)$ that are zero for negative x. One could say that the discrete form of the cosine transform is invertible, but only when attention is restricted to functions $f(\tau)$ that are even [in the sense that $f(\tau)=f(N-\tau)$ for $\tau = 1$ to $N-1$].

Boundary value problems. Useful invertible discrete sine and cosine transforms are generated if in place of frequencies $0, 1/N, 2/N, 3/N \ \dots$, as above, we change to $0, \frac{1}{2}/N, 1/N, 1\frac{1}{2}/N, \dots$. The fundamental frequency is now $1/2N$ and the fundamental period is $2N$. Consider the discrete sum

$$\sum_{\tau=1}^{N-1} f(\tau)\sin(\pi\nu\tau/N).$$

For this operation, attention is restricted to values of τ from 1 to $N-1$; if a value is assigned to $f(0)$, the value will leave no trace, even if the sum $\sum_{\tau=0}^{N-1}$ is

evaluated, because the basis functions $\sin(\pi\nu\tau/N)$ all have a null at $\tau = 0$. This operation, if applied twice in succession, returns $\frac{1}{2}N\, f(\tau)$. Thus we can define a discrete sine transform

$$F_s(\nu) = \frac{2}{N}\sum_{\tau=1}^{N-1} f(\tau)\sin(\pi\nu\tau/N)$$

with an inverse

$$f(\tau) = \sum_{\nu=1}^{N-1} F_s(\nu)\sin(\pi\nu\tau/N).$$

The corresponding discrete cosine transform $F_c(\nu)$ is

$$F_c(\nu) = \frac{2}{N}\sum_{\tau=0}^{N-1} f(\tau)\cos(\pi\nu\tau/N).$$

However, to recover $f(\tau)$, given $F_c(\nu)$, is algebraically complicated. A fast algorithm COSFT, that depends on the FFT, appears in Press et al. (1986). A revised definition introduced in the second edition (1992) based on $N + 1$ data values is strictly symmetrical:

$${}^0F(\nu) = \sqrt{\frac{2}{N}}\left\{\tfrac{1}{2}[f(0) + (-1)^\nu f(N)] + \sum_{\tau=0}^{N-1} f(\tau)\cos(\pi\nu\tau/N)\right\}.$$

A related DCT (Rao and Yip, 1990) departs from the smooth cosinusoidal basis functions used above, for example by substituting $1/\sqrt{2}$ for unity where the argument of cosine is zero. Thus

$${}^{\mathrm{I}}F_c(\nu) = \sqrt{\frac{2}{N}}k_\nu \sum_{\tau=0}^{N} k_\tau f(\tau)\cos(\pi\nu\tau/N), \qquad \nu = 0 \text{ to } N.$$

The factors k_ν and k_τ are unity, except when the subscript is zero or N, in which case the value is $1/\sqrt{2}$.

Whereas hitherto N was the number of elements comprising $f(\tau)$, here it is the number of elements minus 1 (there are $N + 1$ elements). This transform is strictly symmetrical:

$$f(\tau) = \sqrt{\frac{2}{N}}k_\tau \sum_{\nu=0}^{N} k_\nu\, {}^{\mathrm{I}}F_c(\nu)\cos(\pi\nu\tau/N), \qquad \nu = 0 \text{ to } N,$$

and is known as DCT1.

A body of lore for the solution of differential equations without involving complex Fourier analysis is responsible for the importance of the discrete sine and cosine transforms. As a simple example, if a violin string of length L is drawn aside at a noncentral point and one wishes to study the consequences, it is not necessary to apply full Fourier analysis of the string shape $y(x)$. Components restricted to the series $\sin(\pi x/L)$, $\sin(2\pi x/L)$, $\sin(3\pi x/L)$, and so on, suffice to syn-

thesize the string shape because the boundary conditions $y(0) = 0$ and $y(L) = 0$ render cosine terms unnecessary. In another problem where, instead of $y(0) = 0$ the boundary condition is that $[dy/dx]_{x=0,L}$ equals zero, cosine components alone may suffice. While it is certainly possible to do these boundary value problems with the full Fourier transform, it is not customary. Normal Fourier analysis would yield both sine and cosine components instead of just one kind alone. Thus, taking $L = 1$ and

$$f(x) = \begin{cases} 2x & 0 < x < a \\ 1 - x & a < x < 1 \\ 0 & \text{elsewhere} \end{cases}$$

then the complex Fourier transform is

$$F(s) = \frac{1}{2}\operatorname{sinc}^2 s - \frac{1}{2}a \operatorname{sinc}^2 as + i\frac{\sin 2\pi s - 3 \sin 2\pi as}{4\pi^2 s^2}.$$

Inverting, $f(x)$ may be expressed as $\int_{-\infty}^{\infty} F(s)e^{i2\pi sx}ds$, where the integrand is complex.

But for the range $0 \leqslant x \leqslant \infty$, we can also write the simpler relation

$$f(x) = 2\int_0^\infty \frac{\sin 2\pi s - 3\sin 2\pi as}{4\pi^2 s^2} \sin 2\pi sx\, ds.$$

where the integrand is purely real, consisting only of sine waves of different amplitudes and frequencies.

No matter how complicated the initial shape of the string may be at the moment of release, and whether or not any symmetry or antisymmetry is present, it is striking that the string shape can be analyzed into sine components alone. Waves in a canal with closed ends can be analyzed into cosine components alone.

Fields of physics involving the wave equation, Laplace's equation, the diffusion equation, and other fundamental differential equations in two, three, and four dimensions, provide a wide range of application for numerical analysis using the discrete sine and cosine transforms.

Data compression application. When images have to be transmitted digitally computers are generally not fast enough to preserve photographic quality, but much successful effort has gone into reducing the number of bits transmitted while simultaneously minimizing deterioration of the received image. If an image represented by a million samples, each sample quantized to 256 levels, has too many bits for some purpose, then a fourfold reduction in the number of bits will result if sets of 2×2 pixels are replaced by one pixel, to which is assigned the mean value of the pixels replaced. The deterioration will consist of blurring describable as convolution followed by coarser sampling. If the image is printed with typical newspaper quality, spread over an area about the size of a sheet of letter paper, the blurring will be distinctly noticeable to the eye. There will be an additional 32-fold reduction in the number of bits because newsprint offers only about eight grey levels,

but for many purposes, such as recognizing a face, this would be acceptable deterioration. It is clear that the criteria of acceptability will involve the psychophysics of vision and the character of the image in addition to the digitizing procedure.

Empirical study shows that many images when subjected to spatial harmonic analysis have little content at high spatial frequencies. This suggests transmitting the Fourier transform of the image, but with the higher frequencies suppressed or deemphasized (for example by quantizing to fewer levels). At the receiving end, inverting the transform will yield an image whose deterioration may be acceptable. This procedure is known as transform encoding. Whether such data compression is implemented in the image domain or in the transform domain, all parts of the image are treated equally; still, it is apparent that blurring is less acceptable in some areas of an image than in others: on a page of text for example. Therefore, further gain can be achieved by one-dimensional data manipulation such as run-length encoding, TIFF encoding, and delta-row encoding. Segmentation of an image into parts characterized by more or less uniform spatial statistics offers further opportunities.

When statistical properties of images are taken into account, data-compression advantages may be claimed on behalf of one transform over another. Such a preferred transform, and its inverse, are defined by

$$ {}^{\mathrm{II}}F(\nu) = \frac{\alpha(\nu)}{N} \sum_{\tau=0}^{N-1} f(\tau) \cos\left[\frac{\pi}{2N}(2\tau + 1)\nu\right] $$

$$ f(\tau) = \sum_{\nu=0}^{N-1} \alpha(\nu)\,{}^{\mathrm{II}}F(\nu) \cos\left[\frac{\pi}{2N}(2\tau + 1)\nu\right], $$

where $\alpha(\nu) = 1$ except that $\alpha(0) = 1/\sqrt{2}$. The transform $F(\nu)$ was introduced as "a DCT" by Ahmed, Natarajan, and Rao (1974). This influential letter ultimately led to mention of this DCT in an international standard described in Pennebaker and Mitchell (1993). It is known as DCT2 (Rao and Yip, 1990).

The DCT2 analyzes into frequencies $0, \frac{1}{2}/N, 1/N, 1\frac{1}{2}/N, \cdots$ just as the DCT1 does, but the samples are time shifted by half a unit and the basis functions are not exactly cosinusoidal. Thus the cosine factor is no longer unity at $\tau = 0$ but has a variable value $\cos(\pi\nu/2N)$. Nevertheless, the basis vectors are orthogonal. The transformation is not strictly invertible; that is, if applied twice in succession, it does not return the original object. Familiar theorems for the DFT and the Hartley transform, the shift and convolution theorems, for example, become opaque. For computation it is found convenient, as exemplified by Press et al. and MATLAB (which offers an operation `dct(f)`), to rearrange the data so that the FFT can be applied. Alternatively see code that has been published by several authors to apply the Hartley transform to arrive at DCT2 (p. 65 of Rao and Yip, 1990). The main distinguishing advantage claimed for the DCT2 is favorable performance as judged empirically on images whose statistics approximate spatially uniform first order Markov statistics. Transmission of text images may not benefit much from data compression nor will time signals such as ASCII or cellular telephone transmissions, or digitized speech or music.

COMPUTING

The properties of the discrete Hartley transform commend themselves for application to numerical analysis. Where speed is of the essence one may use the "fast Hartley transform" (FHT) described next. Many users' programs will run significantly faster with a fast Hartley transform than with the FFT; and on personal computers the simplicity of the Hartley transform is an advantage. The practice of making up pseudo-complex data from real data to suit the character of the Fourier transform has been obsoleted by the introduction of means for operating directly on real data (Buneman, 1986).

A theoretical convenience of the Fourier transform is that its complex kernel is an eigenfunction of a linear time-invariant operator. This algebraic property is valueless in the computing environment, neither the sine, cosine, nor cas function being an eigenfunction.

GETTING A FEEL FOR NUMERICAL TRANSFORMS

Experience with Fourier integral transform pairs presented graphically, as in the Pictorial Dictionary, develops a feeling for the interpretation of frequency analysis that transfers readily to the corresponding numerical transforms. Of course, the practice of avoiding negative indices, which is inherited from early programmers, clashes with the idea of negative frequency that is imbedded in mathematical analysis as exemplified by

$$\cos \omega t = \frac{1}{2}e^{-i\omega t} + \frac{1}{2}e^{+i\omega t}.$$

The analysis of a cosine function into two frequencies, one negative and one positive, each with amplitude $\frac{1}{2}$, is reinforced by that familiar diagram where there is a cosine on the left and two half-strength impulses on the right (Fig. 6.1). If $\cos(2\pi t/4)$, which has period 4, is discretized at unit interval of t, one would connect that diagram with the discretized version if the pair were presented numerically as

$$\begin{array}{ccc} \tau = 0 & & \nu = -2 \qquad \nu = +2 \\ \ldots 0 \quad -1 \quad 0 \;\; \mathbf{1} \;\; 0 \quad -1 \quad 0 \ldots & \qquad & \ldots 0 \;\; -\tfrac{1}{2} \;\; 0 \quad \mathbf{0} \quad +\tfrac{1}{2} \;\; 0 \;\; 0 \ldots \\ \text{waveform} & & \text{spectrum} \end{array}$$

For $\sin(2\pi t/4)$ we would also recognize

$$\begin{array}{ccc} \tau = 0 & & \nu = 0 \\ \ldots 1 \;\; 0 \quad -1 \;\; \mathbf{0} \;\; 1 \;\; 0 \quad -1 \ldots & \qquad & \ldots 0 \;\; 0 \;\; +\tfrac{1}{2} \;\; 0 \;\; \mathbf{0} \;\; 0 \;\; -\tfrac{1}{2} \;\; 0 \;\; 0 \ldots \\ \text{waveform} & & \text{spectrum} \end{array}$$

and if some d.c. were added, say of strength 2, the graphical form of the statement

$$\cos(2\pi t/4) \supset \tfrac{1}{2}\delta(t+\tfrac{1}{4}) + 2\delta(t) + \tfrac{1}{2}\delta(t-\tfrac{1}{4})$$

would connect with

$$\begin{array}{cc} \tau = 0 & \nu = 0 \\ \ldots 2\ 1\ 2\ \mathbf{3}\ 2\ 1\ 2 \ldots & \ldots 0\ 0\ +\tfrac{1}{2}\ 0\ \mathbf{2}\ 0\ +\tfrac{1}{2}\ 0\ 0 \ldots. \\ \text{waveform} & \text{spectrum} \end{array}$$

The origins of time ($\tau = 0$) and of frequency ($\nu = 0$) have been emphasized by bold type.

However, the indexing convention presents us with the following for the three examples above.

$$\begin{array}{l} \tau = 0 \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \nu = 2 \qquad \nu = 6 \\ \{\mathbf{1}\ \ 0\ \ -1\ \ \ 0\ \ 1\ \ 0\ \ -1\ \ \ 0\} \supset \{\mathbf{0}\ \ 0\ \ \ \tfrac{1}{2}\ \ 0\ \ 0\ \ 0\ \ \tfrac{1}{2}\ \ 0\} \\ \{\mathbf{0}\ \ 1\ \ \ 0\ \ -1\ \ 0\ \ 1\ \ \ 0\ \ -1\} \supset \{\mathbf{0}\ \ 0\ \ -\tfrac{1}{2}\ \ 0\ \ 0\ \ 0\ \ \tfrac{1}{2}\ \ 0\} \\ \{\mathbf{3}\ \ 2\ \ \ 1\ \ \ 2\ \ 3\ \ 2\ \ \ 1\ \ \ 2\} \supset \{\mathbf{2}\ \ 0\ \ \ \tfrac{1}{2}\ \ 0\ \ 0\ \ 0\ \ \tfrac{1}{2}\ \ 0\} \end{array}$$

The frequencies 1/8, 2/8, 3/8, and 4/8 are represented by ν-values of 1, 2, 3, and 4, while frequencies $-1/8$, $-2/8$, $-3/8$ are represented by ν-values of 7, 6, 5 respectively.

We recall that ν/N gives the frequency, for nonnegative frequencies ($0 \leqslant \nu \leqslant N$), while a negative frequency such as -0.25, which could have fallen at $\nu = -2$ if negative indices were allowed, is placed at $-2 \bmod 8$, i.e., at $\nu = 6$. When we look at a numerical spectrum, the first element gives the d.c. value (just as the leading Fourier series coefficient a_0 does), while a frequency f is represented at the two locations $\nu = (\pm Nf) \bmod N$. Cosine components have equal values at these locations; sine components are equal with opposite signs.

All of the above examples apply to the discrete Fourier transform and to the discrete Hartley transform. For the other real valued transforms there is more to be said. For comparison, here are the three transforms presented above for both the DCT1 (middle) and the DCT2 (right).

$f(\tau)$	DCT1	DCT2
{**1** 0 −1 0 1 0 −1 0}	{**.11** .47 .32 1.31 1.31 −.32 .41 −.11}	{**0** .53 0 1.09 1.41 −.73 0 −.11}
{**0** 1 0 −1 0 1 0 −1}	{**0** 1 0 −1 0 1 0 −1}	{**5.66** .53 0 1.09 1.41 −.73 0 −.11}
{**3** 2 1 2 3 2 1 2}	{**5.49** .41 .12 1.31 1.75 −.32 .85 −.11}	{**0** .53 0 1.09 −1.41 0 −.73 −.11}

THE COMPLEX HARTLEY TRANSFORM

Since the radiation pattern of an antenna is expressible as the Fourier transform of an aperture distribution of an electromagnetic field, and since a lens can con-

vert an optical aperture distribution to another distribution that is similarly expressible, it might appear that the Hartley transform is a mathematical construct without physical interpretation. However, an intimate relationship exists between the two transforms such that, whatever property one of the transforms may possess, a corresponding property is possessed by the other. This relationship can be explained with the aid of a diagram introduced in Fig. 2.10, where the real and imaginary parts of a Fourier transform $F(s)$ are plotted as a locus on the complex plane of $R = \mathrm{Im}\, F$ versus $S = \mathrm{Re}\, F$, each point on the locus parameterized by a value of s. If a second coordinate system R',S' is constructed by rotating the R- and S-axes through $-45°$ and the component in the R' direction is read off as a function of s, the result will be the Hartley transform. If on the other hand the component in the S' direction is examined it will be found to add no information, in the case of the transform of a real function $f(x)$. The transformation

$$R'(s) + iI'(s) = [R(s) + iI(s)] \times e^{i\pi/4}$$

describes the intimate relationship between the two transforms. The complex-plane locus can be reconstructed from $R(s)$ and $I(s)$ together; but the absence of the redundancy associated with the symmetries of $R(s)$ and $I(s)$ permits the real $f(x)$ to be recovered from the component in the R'-direction alone, or indeed from any direction save only the R- and S-directions.

For an original function $f(x)$ representing *complex* data, the S'-direction handles the additional information; the Hartley transform, as conventionally defined, then becomes complex. Millane (1994) gives an extended treatment of the complex Hartley transform.

PHYSICAL ASPECTS OF THE HARTLEY TRANSFORMATION

From this complex-plane view the construction of the Hartley transform in physical space can be derived (Bracewell, 1989). The first apparatus to demonstrate the possibility of phase recovery from a real electromagnetic field distribution utilized an optical beamsplitter (Bracewell et al., 1985; Bracewell, 1986; Villasenor and Bracewell, 1987; Villasenor, 1989) and Michelson interferometer technique. Later, appropriate substitution of microwave elements enabled the wavelength range of feasibility to be extended (Villasenor and Bracewell, 1988).

In the Hartley plane, phase is encoded by amplitude alone; in fact when the apparatus is in proper adjustment the plane is in isophase, except for the phase reversal where the mathematical Hartley transform would be negative. The possibility of obtaining electromagnetic phase by amplitude or intensity measurement alone is of technical interest because at ultraviolet and x-ray wavelengths phase is important but difficult to measure. Implementation with x-rays has been discussed but will be technically difficult and needs further consideration.

THE FAST HARTLEY TRANSFORM

A sequence of N real numbers possesses a discrete Hartley transform (DHT) that is a sequence of the same length and is also real valued. From the DHT one can return to the original sequence by applying the same transformation formula a second time. The convenience of not having to manage the real and imaginary parts either in separate arrays, or interleaved in one array of double length, or in other ingenious ways that have been adopted in various embodiments of the Fourier transform commends the DHT for consideration in applications to numerical spectral analysis and convolution. Not having to allow for an inverse transformation that is different and differs among authors is also helpful.

The DHT is a suitable *substitute* for the discrete Fourier transform (DFT) for some purposes; however, if the real and imaginary parts of the DFT are expressly required, then they are directly obtainable as the even and odd parts of the DHT. When the power spectrum is the desired goal, it may be obtained directly from the DHT without first calculating the real and imaginary parts of the DFT as in the usual way of calculating power spectra.

A fast algorithm has been developed for computing the DHT and, by analogy, with the fast Fourier transform will be referred to as the fast Hartley transform (FHT). A subprogram appears in an explanatory context in Bracewell (1986) under the name FHTSUB. At the end of this chapter a condensed radix-4 version appears. In order to make the material as accessible as possible to personal-computer users the program is presented in pseudocode that is as free from idiosyncrasies as possible and ready to be transcribed into any current computer language.

In addition to the convenient features mentioned, the FHT is faster than the complex FFT by a factor that depends on circumstances discussed below. There are one-way programs related to the FFT that are available for operating only on real data and save time by omitting the computation of those coefficients that are complex conjugates of other coefficients. Nevertheless, the real input leads to complex output; to invert the complex output requires storage of an additional one-way program which is suitably modified to accept complex data and which saves time by not computing imaginary parts. By comparison the FHT, which is bidirectional and real-valued, is simpler and more elegant. Neither of the fast one-way programs can accept its own output. For later developments see Guo et al. (1998).

When the fast Hartley algorithm appeared in 1984 some incredulity was expressed about the merits of speed, reality, and reversibility claimed for it by comparison with the familiar FFT, but hundreds of papers appeared (Olejniczak and Heydt, 1994), chips were manufactured for incorporation into commercial instruments, and the physical significance of the Hartley transform was demonstrated in the laboratory with microwaves and visible light. Engineering applications answered a criticism that the Fourier transform was a property of nature whereas the Hartley was not, a difficult claim to defend, seeing that complex num-

bers are a mental construct. It soon became clear that the relationship is very intimate, amounting to full equivalency.

We recall that for a real function $f(\tau)$, for $\tau = 0, 1, \ldots, N-1$, one defines the discrete Hartley transform $H(\nu)$ by

$$H(\nu) = N^{-1} \sum_{\tau=0}^{N-1} f(\tau) \operatorname{cas}\left(\frac{2\pi\nu\tau}{N}\right), \qquad \nu = 0, 1, \ldots, N-1.$$

The inverse relation is

$$f(\tau) = \sum_{\nu=0}^{N-1} H(\nu) \operatorname{cas}\left(\frac{2\pi\nu\tau}{N}\right), \qquad \tau = 0, 1, \ldots, N-1.$$

The integer τ can be thought of as a mnemonic for time, while ν/N, where $\nu \leqslant N/2$, is like frequency measured in cycles per unit of time.

THE FAST ALGORITHM

If the DHT is evaluated numerically from its defining expression, the time taken for long data sequences of length N is proportional to N^2 as with the DFT, because for each value of ν there are N evaluations of the product $f(\tau)\operatorname{cas}(2\pi\nu\tau/N)$ and there are N values of ν. Let N be expressible as 2 raised to some power P, i.e., $N = 2^P$. Just as with the FFT, the number of arithmetic operations can be reduced to the order of $N \log_2 N$ or NP.

One way of proceeding is shown in the flow diagram of Fig. 12.1 for the case of $N = 8$, $P = 3$. The operation labeled PERMUTE, discussed further in a later section, rearranges the sequence of data. The ith member is placed into the jth position where j may be calculated from i as follows.

```
R = 1
J = 0
FOR K = 1 TO P
    S = R DIV 2
    J = J + J + R - S - S
    R = S
NEXT K
```

The purpose of permutation is the same as with the FFT, namely, to bisect the data sequence progressively until data pairs are reached. By definition when $N = 2$,

$$\{a \quad b\} \quad \text{has DHT} \quad \tfrac{1}{2}\{a+b \quad a-b\}$$

which is very simple. To superimpose all the two-element transforms requires a decomposition formula that expresses the DFT of a given sequence in terms of the DHTs of subsequences of half length. This formula permits, for example, the DHT of a four-element sequence $\{a_1\, a_2\, b_1\, b_2\}$ to be expressed in terms of the DHTs

of the two interleaved two-element sequences $\{a_1\, b_1\}$ and $\{a_2\, b_2\}$. To derive the decomposition formula, we require two theorems, the shift theorem and the similarity theorem. The shift theorem states that if $f(\tau)$ has DHT $H(\nu)$, then

$$f(\tau + a) \quad \text{has DHT} \quad H(\nu)\cos\left(\frac{2\pi a\nu}{N}\right) - \mathrm{H}(\mathrm{N} - \nu)\sin\left(\frac{2\pi a\nu}{N}\right).$$

The similarity theorem, which is the same as for the DFT, states that if a sequence $f(\tau)$ is stretched to double its length by inserting a zero element after each given element, then the elements of the original DHT are repeated. As examples,

$$\{1\,2\,3\,4\} \quad \text{has DHT} \quad \{2.5\ {-1}\ {-0.5}\ 0\},$$

$$\{1\,0\,2\,0\,3\,0\,4\,0\} \quad \text{has DHT} \quad \{2.5\ {-1}\ {-0.5}\ 0\ 2.5\ {-1}\ {-0.5}\ 0\}.$$

Both these theorems are derivable in one line from the DHT definition. Suppose that

$$\{a_1\, a_2\, b_1\, b_2\, c_1\, c_2 \ldots\} \quad \text{has DHT} \quad H(\nu)$$

where there are N elements. Then the a_1 sequence of $\frac{1}{2}N$ elements $\{a_1\, b_1\, c_1 \ldots\}$ has DHT $\{\alpha_1\, \beta_1\, \gamma_1 \ldots\}$ and the a_2 sequence $\{a_2\, b_2\, c_2 \ldots\}$ has DHT $\{\alpha_2\, \beta_2\, \gamma_2 \ldots\}$. We see that

$$\begin{aligned}
H(\nu) &= \text{DHT of } \{a_1\, 0\, b_1\, 0\, c_1\, 0 \ldots\} + \text{DHT of } \{0\, a_2\, 0\, b_2\, 0\, c_2\, 0 \ldots\} \\
&= \{\alpha_1\, \beta_1\, \gamma_1 \ldots \alpha_1\, \beta_1\, \gamma_1 \ldots\} + \text{DHT of } \{0\, 1\} * \{a_2\, b_2\, c_2 \ldots\} \\
&= \{\alpha_1\, \beta_1\, \gamma_1 \ldots \alpha_1\, \beta_1\, \gamma_1 \ldots\} + \Bigg\{\alpha_2 \quad \beta_2 \cos\left(\frac{2\pi}{N}\right) \quad \gamma_2 \cos\left(\frac{2\pi.2}{N}\right) \ldots \\
&\quad -\alpha_2 \quad \beta_2 \cos\left[\frac{2\pi(N/2 + 1)}{N}\right] \quad \gamma_2 \cos\left[\frac{2\pi(N/2 + 2)}{N}\right] \ldots \Bigg\} \\
&\quad + \Bigg\{0 \ldots \gamma_2 \sin\left[\frac{2\pi(N/2 - 2)}{N}\right] \quad \beta_2 \sin\left[\frac{2\pi(N/2 - 1)}{N}\right] \quad 0 \ldots \\
&\quad \gamma_2 \sin\left[\frac{2\pi(N/2 + 2)}{N}\right] \quad \beta_2 \sin\left[\frac{2\pi(N/2 + 3)}{N}\right]\Bigg\}.
\end{aligned}$$

The general decomposition formula is thus

$$H(\nu) = H_{a_1}(\nu) + H_{a_2}(\nu)\cos\left(\frac{2\pi\nu}{N}\right) + \mathrm{H}_{a_2}(\mathrm{N} - \nu)\sin\left(\frac{2\pi\nu}{N}\right)$$

where $H_{a_1}(\nu)$ and $H_{a_2}(\nu)$ are $\frac{1}{2}\{\alpha_1\, \beta_1\, \gamma_1 \cdots \alpha_1\, \beta_1\, \gamma_1 \cdots\}$ and $\frac{1}{2}\{\alpha_2\, \beta_2\, \gamma_2 \cdots \alpha_2\, \beta_2\, \gamma_2 \cdots\}$, respectively. In this derivation we have put $a = -1$ in the shift theorem. In the first two operations following permutation on the left of Fig. 12.4, the sine and cosine factors assume only values of 0, 1, or -1. For this reason the diagram can make use of two flow line types: unbroken lines that transmit values unchanged and broken lines that transmit with a sign reversal. In later stages sine and cosine factors are associated with two-thirds of the flow lines; of course,

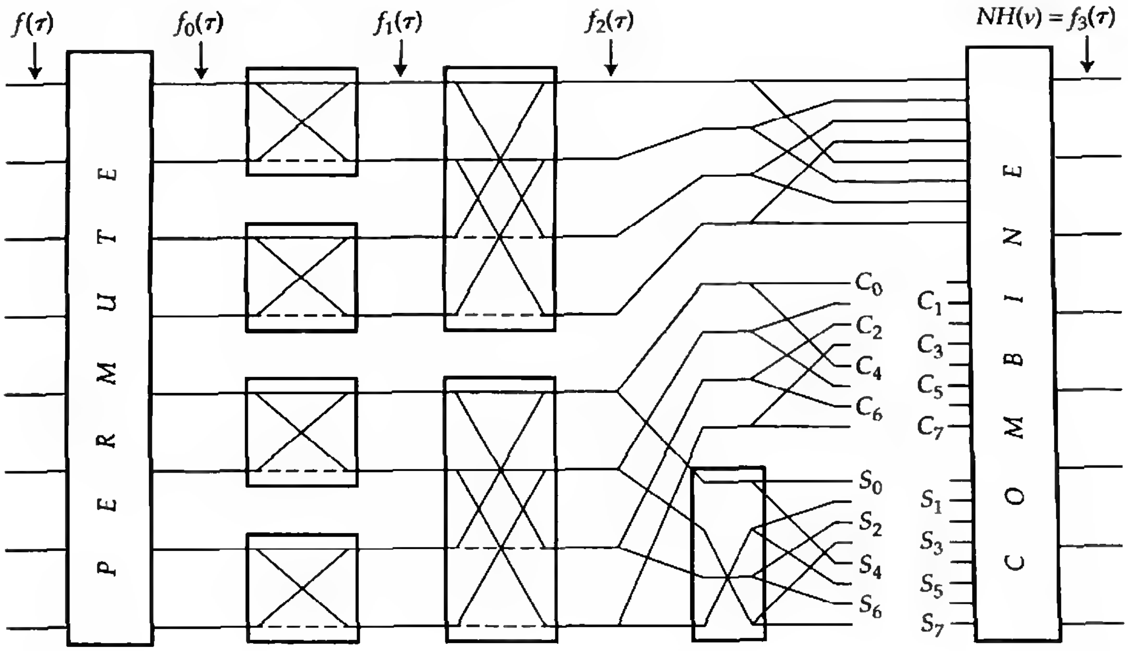


Fig. 12.4 **Flow diagram for the discrete Hartley transform with $N = 8$ and $P = 3$. Broken lines represent transfer factors -1 while full lines represent unity transfer factors. The crossover boxes perform the sign reversal called for by the shift theorem which also requires the sine and cosine factors S_n, C_n.**

many of these factors are still 0, 1, or -1. For the case of $N = 16$, Table 12.1 summarizes the equations represented by the flow diagram using the level-dependent abbreviations C_n and S_n to stand for $^{\cos}_{\sin}(2\pi n/2^L)$, respectively, where L is the level number of the stage. Table 12.2 shows how the equations in fact simplify when the special values 0, 1, -1, and $r = 2^{-1/2}$ are substituted. The small boxes containing crossover connections that appear in one-third of the connections in the later stages (Fig. 12.4) are to implement the sign reversal demanded by the factor $H(N - \nu)$ in the shift theorem. [When $\nu = 0$, $H(N)$ is assigned the value $H(0)$.]

An example (Table 12.3) will clarify the steps involved. As will be seen, the computations for $N = 8$ can easily be carried out by hand. For illustration let the given data sequence be $f(\tau) = \{1\,2\,3\,4\,5\,6\,7\,8\}$. Permutation proceeds in $P - 1$ steps. The first step is to separate out the data into two four-element sequences $\{1\,3\,5\,7\}$ and $\{2\,4\,6\,8\}$, as in the column headed π. The second step separates each four-element sequence into two two-element sequences: $\{1\,5\}$, $\{3\,7\}$, $\{2\,6\}$, and $\{4\,8\}$. As $P = 3$, the second step is the last step of permutation.

The sequence $\{1\,5\,3\,7\,2\,6\,4\,8\}$ is the permuted sequence $f_0(\tau)$. Each two-element sequence $\{a\,b\}$ is now transformed; thus $\{1\,5\}$ gives $\frac{1}{2}\{6\;-4\}$, but we suppress the factors $\frac{1}{2}$ until the end. When concatenated, these elementary transforms

TABLE 12.1

Equations relating successive stages of the FHT as indexed by level number L

Data	Permute	Level 1	Level 2	Level 3	16 × DHT
$F(0,0)$	$F(0,0) = F(0,0)$	$F(1,0) = F(0,0) + F(0,1)$	$F(2,0) = F(1,0) + F(1,2)C_0 + F(1,2)S_0$	$F(3,0) = F(2,0) + F(2,4)C_0 + F(2,4)S_0$	$F(4,0) = F(3,0) + F(3,8)C_0 + F(3,8)S_0$
$F(0,1)$	$F(0,1) = F(0,8)$	$F(1,1) = F(0,0) - F(0,1)$	$F(2,1) = F(1,1) + F(1,3)C_1 + F(1,3)S_1$	$F(3,1) = F(2,1) + F(2,5)C_1 + F(2,7)S_1$	$F(4,1) = F(3,1) + F(3,9)C_1 + F(3,15)S_1$
$F(0,2)$	$F(0,2) = F(0,4)$	$F(1,2) = F(0,2) + F(0,3)$	$F(2,2) = F(1,0) + F(1,2)C_2 + F(1,2)S_2$	$F(3,2) = F(2,2) + F(2,6)C_2 + F(2,6)S_2$	$F(4,2) = F(3,2) + F(3,10)C_2 + F(3,14)S_2$
$F(0,3)$	$F(0,3) = F(0,12)$	$F(1,3) = F(0,2) - F(0,3)$	$F(2,3) = F(1,1) + F(1,3)C_3 + F(1,3)S_3$	$F(3,3) = F(2,3) + F(2,7)C_3 + F(2,5)S_3$	$F(4,3) = F(3,3) + F(3,11)C_3 + F(3,13)S_3$
$F(0,4)$	$F(0,4) = F(0,2)$	$F(1,4) = F(0,4) + F(0,5)$	$F(2,4) = F(1,4) + F(1,6)C_0 + F(1,6)S_0$	$F(3,4) = F(2,0) + F(2,4)C_4 + F(2,4)S_4$	$F(4,4) = F(3,4) + F(3,12)C_4 + F(3,12)S_4$
$F(0,5)$	$F(0,5) = F(0,10)$	$F(1,5) = F(0,4) - F(0,5)$	$F(2,5) = F(1,5) + F(1,7)C_1 + F(1,7)S_1$	$F(3,5) = F(2,1) + F(2,5)C_5 + F(2,7)S_5$	$F(4,5) = F(3,5) + F(3,13)C_5 + F(3,11)S_5$
$F(0,6)$	$F(0,6) = F(0,6)$	$F(1,6) = F(0,6) + F(0,7)$	$F(2,6) = F(1,4) + F(1,6)C_2 + F(1,6)S_2$	$F(3,6) = F(2,2) + F(2,6)C_6 + F(2,6)S_6$	$F(4,6) = F(3,6) + F(3,14)C_6 + F(3,10)S_6$
$F(0,7)$	$F(0,7) = F(0,14)$	$F(1,7) = F(0,6) - F(0,7)$	$F(2,7) = F(1,5) + F(1,7)C_3 + F(1,7)S_3$	$F(3,7) = F(2,3) + F(2,7)C_7 + F(2,5)S_7$	$F(4,7) = F(3,7) + F(3,15)C_7 + F(3,9)S_7$
$F(0,8)$	$F(0,8) = F(0,1)$	$F(1,8) = F(0,8) + F(0,9)$	$F(2,8) = F(1,8) + F(1,10)C_0 + F(1,10)S_0$	$F(3,8) = F(2,8) + F(2,12)C_0 + F(2,12)S_0$	$F(4,8) = F(3,0) + F(3,8)C_8 + F(3,8)S_8$
$F(0,9)$	$F(0,9) = F(0,9)$	$F(1,9) = F(0,8) - F(0,9)$	$F(2,9) = F(1,9) + F(1,11)C_1 + F(1,11)S_1$	$F(3,9) = F(2,9) + F(2,13)C_1 + F(2,15)S_1$	$F(4,9) = F(3,1) + F(3,9)C_9 + F(3,15)S_9$
$F(0,10)$	$F(0,10) = F(0,5)$	$F(1,10) = F(0,10) + F(0,11)$	$F(2,10) = F(1,8) + F(1,10)C_2 + F(1,10)S_2$	$F(3,10) = F(2,10) + F(2,14)C_2 + F(2,14)S_2$	$F(4,10) = F(3,2) + F(3,10)C_{10} + F(3,14)S_{10}$
$F(0,11)$	$F(0,11) = F(0,13)$	$F(1,11) = F(0,10) - F(0,11)$	$F(2,11) = F(1,9) + F(1,11)C_3 + F(1,11)S_3$	$F(3,11) = F(2,11) + F(2,15)C_3 + F(2,13)S_3$	$F(4,11) = F(3,3) + F(3,11)C_{11} + F(3,13)S_{11}$
$F(0,12)$	$F(0,12) = F(0,3)$	$F(1,12) = F(0,12) + F(0,13)$	$F(2,12) = F(1,12) + F(1,14)C_0 + F(1,14)S_0$	$F(3,12) = F(2,8) + F(2,12)C_4 + F(2,12)S_4$	$F(4,12) = F(3,4) + F(3,12)C_{12} + F(3,12)S_{12}$
$F(0,13)$	$F(0,13) = F(0,11)$	$F(1,13) = F(0,12) - F(0,13)$	$F(2,13) = F(1,13) + F(1,15)C_1 + F(1,15)S_1$	$F(3,13) = F(2,9) + F(2,13)C_5 + F(2,15)S_5$	$F(4,13) = F(3,5) + F(3,13)C_{13} + F(3,11)S_{13}$
$F(0,14)$	$F(0,14) = F(0,7)$	$F(1,14) = F(0,14) + F(0,15)$	$F(2,14) = F(1,12) + F(1,14)C_2 - F(1,14)S_2$	$F(3,14) = F(2,10) + F(2,14)C_6 + F(2,14)S_6$	$F(4,14) = F(3,6) + F(3,14)C_{14} + F(3,10)S_{14}$
$F(0,15)$	$F(0,15) = F(0,15)$	$F(1,15) = F(0,14) - F(0,15)$	$F(2,15) = F(1,13) + F(1,15)C_3 + F(1,15)S_3$	$F(3,15) = F(2,11) + F(2,15)C_7 + F(2,13)S_7$	$F(4,15) = F(3,7) + F(3,15)C_{15} + F(3,9)S_{15}$

TABLE 12.2

Simplification permitted by explicit substitution for the sine and cosine factors of Table 12.1, many of which are 0, 1, or −1

Data	Permute	Level 1	Level 2	Level 3
$F(0,0)$	$F(0,0) = F(0,0)$	$F(1,0) = F(0,0) + F(0,1)$	$F(2,0) = F(1,0) + F(1,2)$	$F(3,0) = F(2,0) + F(2,4)$
$F(0,1)$	$F(0,1) = F(0,8)$	$F(1,1) = F(0,0) - F(0,1)$	$F(2,1) = F(1,1) + F(1,3)$	$F(3,1) = F(2,1) + rF(2,5) + rF(2,7)$
$F(0,2)$	$F(0,2) = F(0,4)$	$F(1,2) = F(0,2) + F(0,3)$	$F(2,2) = F(1,0) - F(1,2)$	$F(3,2) = F(2,2) \quad + F(2,6)$
$F(0,3)$	$F(0,3) = F(0,12)$	$F(1,3) = F(0,2) - F(0,3)$	$F(2,3) = F(1,1) - F(1,3)$	$F(3,3) = F(2,3) - rF(2,7) + rF(2,5)$
$F(0,4)$	$F(0,4) = F(0,2)$	$F(1,4) = F(0,4) + F(0,5)$	$F(2,4) = F(1,4) + F(1,6)$	$F(3,4) = F(2,0) - F(2,4)$
$F(0,5)$	$F(0,5) = F(0,10)$	$F(1,5) = F(0,4) - F(0,5)$	$F(2,5) = F(1,5) + F(1,7)$	$F(3,5) = F(2,1) - rF(2,5) - rF(2,7)$
$F(0,6)$	$F(0,6) = F(0,6)$	$F(1,6) = F(0,6) + F(0,7)$	$F(2,6) = F(1,4) - F(1,6)$	$F(3,6) = F(2,2) \quad - F(2,6)$
$F(0,7)$	$F(0,7) = F(0,14)$	$F(1,7) = F(0,6) - F(0,7)$	$F(2,7) = F(1,5) - F(1,7)$	$F(3,7) = F(2,3) + rF(2,7) - rF(2,5)$
$F(0,8)$	$F(0,8) = F(0,1)$	$F(1,8) = F(0,8) + F(0,9)$	$F(2,8) - F(1,8) + F(1,10)$	$F(3,8) = F(2,8) + F(2,12)$
$F(0,9)$	$F(0,9) = F(0,9)$	$F(1,9) = F(0,8) - F(0,9)$	$F(2,9) = F(1,9) + F(1,11)$	$F(3,9) = F(2,9) + rF(2,13) + rF(2,15)$
$F(0,10)$	$F(0,10) = F(0,5)$	$F(1,10) = F(0,10) + F(0,11)$	$F(2,10) = F(1,8) - F(1,10)$	$F(3,10) = F(2,10) \quad + F(2,14)$
$F(0,11)$	$F(0,11) = F(0,13)$	$F(1,11) = F(0,10) - F(0,11)$	$F(2,11) = F(1,9) - F(1,11)$	$F(3,11) = F(2,11) - rF(2,15) + rF(2,13)$
$F(0,12)$	$F(0,12) = F(0,3)$	$F(1,12) = F(0,12) + F(0,13)$	$F(2,12) = F(1,12) + F(1,14)$	$F(3,12) = F(2,8) - F(2,12)$
$F(0,13)$	$F(0,13) = F(0,11)$	$F(1,13) = F(0,12) - F(0,13)$	$F(2,13) = F(1,13) + F(1,15)$	$F(3,13) = F(2,9) - rF(2,13) + rF(2,15)$
$F(0,14)$	$F(0,14) = F(0,7)$	$F(1,14) = F(0,14) - F(0,15)$	$F(2,14) = F(1,12) - F(1,14)$	$F(3,14) = F(2,10) \quad - F(2,14)$
$F(0,15)$	$F(0,15) = F(0,15)$	$F(1,15) = F(0,14) - F(0,15)$	$F(2,15) = F(1,13) - F(1,15)$	$F(3,15) = F(2,11) + rF(2,15) - rF(2,13)$

TABLE 12.3
Numerical example of a short FHT with $N = 8, P = 3$

τ	$f(\tau)$	π	$f_0(\tau)$	$f_1(\tau)$	$f_2(\tau)$	$f_3(\tau)$	$H(\nu)$	ν
0	1	1	1	6	16	36	4.5	0
1	2	3	5	−4	−8	−13.6	−1.7	1
2	3	5	3	10	−4	−8	−1	2
3	4	7	7	−4	0	−5.6	−0.7	3
4	5	2	2	8	20	−4	−0.5	4
5	6	4	6	−4	−8	−2.4	−0.3	5
6	7	6	4	12	−4	0	0	6
7	8	8	8	−4	0	5.6	0.7	7

constitute the first stage $f_1(\tau) = \{6\ \ -4\ \ 10\ \ -4\ \ 8\ \ -4\ \ 12\ \ -4\}$. Owing to the degeneracy exhibited in Table 12.1, the steps for obtaining $f_2(\tau)$ are also very simple. Thus $16 = 6 + 10$, $\quad -8 = -4 - 4$, $\quad -4 = 6 - 10$, $\quad 0 = -4 + 4$, etc. Some of the flow lines from Fig. 12.4 are included in Table 12.3 to facilitate cross-reference. In the final stage, computation of $f_3(\tau)$ involves sines and cosines of eighths of a turn. In the operation labeled COMBINE there are three sets of eight inputs each and one set of eight outputs. The first of the eight outputs is the sum of the first elements of the three input sets and similarly for the next outputs.

We remember that a factor $\frac{1}{2}$ was suppressed P times so the result $f_3(\tau)$ must be divided by 8 to conclude that

$$\{1\,2\,3\,4\,5\,6\,7\,8\} \quad \text{has DHT} \quad \{4.5\quad -1.7\quad -1\quad -0.7\quad -0.5\quad -0.30\quad 0.7\}.$$

As a check we note that the sum of the data equals 8 times the first element of the DHT. A study by Hou (1987) led to a different flow diagram with features favorable for VLSI implementation.

If we wished to go on and get the DFT $R(\nu) + jX(\nu)$, we would proceed as follows. The real part $R(\nu)$ is equal to the even part of the DHT $H(\nu)$, and the imaginary part $X(\nu)$ equals the negative odd part. If we want the "power spectrum" $Z^2 = R^2 + X^2$, we do not have to go via the complex quantity $R + jX$; the power spectrum can be calculated directly from

$$Z^2 = \frac{[H(\nu)]^2 + [H(-\nu)]^2}{2}.$$

RUNNING TIME

It is of interest to know how the FHT compares with the FFT in speed. The ratio depends on several factors including the language and the machine; this paper reports on experience with the HP-85 personal computer which provided repeatable run times. In order to normalize the numerical results, one may introduce a figure of merit Y as follows. In the FFT the innermost arithmetic operation is a

complex multiplication which can be carried out by four real multiples, although a very clever method with three multiplies was invented by O. Buneman (1973). One may determine the time T_4 for four real multiples by the following arbitrary but definite prescription:

```
A=RND; B=RND; C=RND; D=RND
T0=TIME
FOR I=1 TO 1000; R=A*C-B*D; X=B*C+A*D; NEXT I
T4=(TIME-T0)/1000; PRINT T4; END
```

Then the running time required to take the FHT of a sequence of length $N = 2^P$ may be expressed as

$$T_{FHT} = YNPT_4.$$

For a standardized comparison I translated a well-known fast Fourier transform program so as to run on my computer (see later section). In the range $1 < P < 11$, Y ranged from 3.1 to 5.8. Using the fast Hartley transform to get the Fourier transform yielded $Y = 1.3$ approximately. Comparison with other programs will require tests in a common language or languages, including assembly language, on a common machine. Programming style, which is important and accounts for some of the discrepancy in the test reported, also needs to be balanced. The fact that Y turns out to depend on N is an interesting reality that is not brought out simply by counting operations in the inner loop and will now be examined.

TIMING VIA THE STRIPE DIAGRAM

Running time dependence on N has been studied using a program given below. For this purpose, and to keep the program as clear as possible, some simplifications have been adopted. For example, each successive stage is allocated to a separate array $F(L, I)$, where L is the "level" or stage number and I ranges from 0 to $N - 1$. In the notation used so far, $F(L, I) \equiv f_L(I)$. However $F(0, I)$ is used first for the data $f(\tau)$ and then reused for the permuted data $f_0(\tau)$. Stages 1 and 2 are then computed directly from the short equations of Table 12.2, without recourse to sines or cosines. After that an iterative loop is entered; as each stage is completed, execution returns to the beginning of the loop.

By varying N we see how the components of Y behave in the "stripe diagram" of Fig. 12.5. There are three kinds of component to be prepared for. First there are some "overhead" operations that do not depend much, if at all, on N and can be expected to die out rapidly when the ordinate is normalized with respect to NP. Second, there are operations such as the precalculation of the needed sines and cosines of multiples of one Nth of a turn that are carried out a number of times proportional to N. (It is desirable to precalculate these factors; otherwise many of them must be redone many times.) Therefore, the contribution of the precalculation to the running time after division by NP dies out as P^{-1} as exhibited

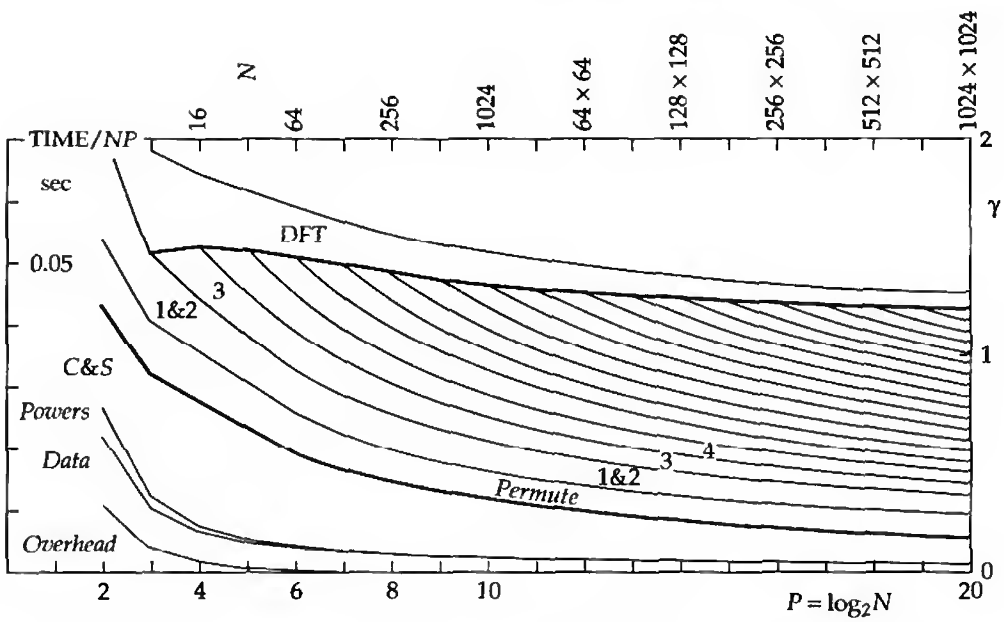


Fig. 12.5 **Stripe diagram showing the figure of merit Y versus P (upper heavy curve) for the FHT program printed below and contributions made by overhead, data insertion, precalculation of powers of 2 and sines and cosines, permutation, and the various numbered stages. The additional time involved in going on to the real and imaginary parts of the discrete Fourier transform is shown by the top stripe. The ordinate scale on the left, which is specific to the author's personal computer, shows running time divided by *NP*.**

by the declining vertical thickness of the stripe labeled "C & S." A second example is the contribution of, let us say, stage 3. As N becomes larger, stage 3 requires more time in proportion to N. Therefore the thickness of the stripe labeled "3" also decays as P^{-1}. But, the number of such stripes increases as P; the zone bounded by heavy lines remains of approximately constant vertical thickness.

The third kind of behavior is exhibited by the contribution from permutation. Permutation involves N reassignments of variables, repeated P times; so the time required is proportional to NP and the permutation stripe therefore does not die out. Any effort made to streamline permutation is consequently most valuable. Merely replacing $J/2$ by $0.5 * J$ or $2 * J$ by $J + J$ may make a significant improvement in the end.

Stages 1 and 2 have been lumped together for the purposes of the stripe diagram. It is apparent that the stripe labeled "1 & 2" is of much the same width as the stripe for stage 3. When stages 1 and 2 were computed by the same general equations as for later stages there was a loss of speed that was quantified by timing and found to be significant. This is the justification for using the short equations of Table 12.2. Further, but diminishing, gains would also be realized from further postponement of iteration.

If the final step to the DFT is taken, a thin decaying band of the second kind is added on top, ultimately becoming negligible.

The nondecaying zone, which is indicated by heavy lines in the stripe diagram, has a thickness of 0.038 s. This means that the running time, in the limit as P approaches infinity, will be $0.038NP$. In the neighborhood of $P = 10$ the running time is about one-third more than this. By special tricks one might hope to erode this one-third. The widest component comes from pretabulation using built-in sine and cosine functions and can be substantially reduced by an algorithm due to O. Buneman (1986) or, as he explains, eliminated entirely when runs are repeated or dedicated hardware is introduced. Overhead and precalculation of powers of 2 are both negligible. Fetching and loading data are shown by this analysis to be also negligible in the current implementation. Permutation is possibly near the ideal limit.

MATRIX FORMULATION

A different and condensed view of the fast Hartley operator may be gained by formulating the equations of Table 12.1 in matrix form. We may write

$$\mathbf{H} = N^{-1}\mathbf{L}_4\mathbf{L}_3\mathbf{L}_2\mathbf{L}_1\mathbf{P}\mathbf{f},$$

where $\mathbf{f}$ and $\mathbf{H}$ are, respectively, the N-element column matrices representing the data $f(\tau)$ and the discrete Hartley transform $H(\nu)$, $\mathbf{P}$ is the permutation matrix and the $\mathbf{L}_i$ are matrix operators which convert the column matrix operand to the column matrix of level i. In this example with $N = 16$ and $P = 4$, i runs from 1 to 4. A subsequent step of conversion to the discrete Fourier transform $F(\nu)$ may also be represented by the (complex) matrix multiplication

$$\mathbf{F} = \mathbf{\Phi}\mathbf{H}.$$

Combining the above equations, we obtain a new expression of the discrete Fourier transform:

$$\mathbf{F} = N^{-1}\mathbf{\Phi}\mathbf{L}_P\mathbf{L}_{P-1}\cdots\mathbf{L}_1\mathbf{P}\mathbf{f}.$$

The matrix operator $N^{-1}\mathbf{\Phi}\mathbf{L}_P \ldots \mathbf{L}_1\mathbf{P}$ thus represents a new factorization of the DFT matrix operator $\mathbf{W}$, where $\mathbf{F} = \mathbf{W}\mathbf{f}$, $W = \exp(-i2\pi/N)$ and

$$\mathbf{W} = \begin{bmatrix} 1 & 1 & 1 & & & \\ 1 & W & W^2 & & & \\ 1 & W^2 & W^4 & & & \\ & & & & & \\ & & & W^{(N-3)^2} & W^{(N-3)(N-2)} & W^{(N-1)(N-3)} \\ & & & W^{(N-2)(N-3)} & W^{(N-2)^2} & W^{(N-2)(N-1)} \\ & & & W^{(N-1)(N-3)} & W^{(N-1)(N-2)} & W^{(N-1)^2} \end{bmatrix}$$

The factors, which are directly verifiable from Table 12.1 are as follows:

$$
\mathbf{P} = \left[\begin{array}{cccc|cccc|cccc|cccc}
1&&& & &&& & &&& & &&& \\
&&& & &&& & 1&&& & &&& \\
&&& & 1&&& & &&& & &&& \\
&&& & &&& & &&& & 1&&& \\
\hline
&&1& & &&& & &&& & &&& \\
&&& & &&& & &&1& & &&& \\
&&& & &&1& & &&& & &&& \\
&&& & &&& & &&& & &&1& \\
\hline
&1&& & &&& & &&& & &&& \\
&&& & &&& & &1&& & &&& \\
&&& & &1&& & &&& & &&& \\
&&& & &&& & &&& & &1&& \\
\hline
&&&1 & &&& & &&& & &&& \\
&&& & &&& & &&&1 & &&& \\
&&& & &&&1 & &&& & &&& \\
&&& & &&& & &&& & &&&1
\end{array}\right]
$$

$$
\mathbf{L}_1 = \left[\begin{array}{cccc|cccc|cccc|cccc}
1&1&& & &&& & &&& & &&& \\
1&-1&& & &&& & &&& & &&& \\
&&1&1 & &&& & &&& & &&& \\
&&1&-1 & &&& & &&& & &&& \\
\hline
&&& & 1&1&& & &&& & &&& \\
&&& & 1&-1&& & &&& & &&& \\
&&& & &&1&1 & &&& & &&& \\
&&& & &&1&-1 & &&& & &&& \\
\hline
&&& & &&& & 1&1&& & &&& \\
&&& & &&& & 1&-1&& & &&& \\
&&& & &&& & &&1&1 & &&& \\
&&& & &&& & &&1&-1 & &&& \\
\hline
&&& & &&& & &&& & 1&1&& \\
&&& & &&& & &&& & 1&-1&& \\
&&& & &&& & &&& & &&1&1 \\
&&& & &&& & &&& & &&1&-1
\end{array}\right]
$$

$$
\mathbf{L}_2 = \left[\begin{array}{cccc|cccc|cccc|cccc}
1&&1& & &&& & &&& & &&& \\
&1&&1 & &&& & &&& & &&& \\
1&&-1& & &&& & &&& & &&& \\
&1&&-1 & &&& & &&& & &&& \\
\hline
&&& & 1&&1& & &&& & &&& \\
&&& & &1&&1 & &&& & &&& \\
&&& & 1&&-1& & &&& & &&& \\
&&& & &1&&-1 & &&& & &&& \\
\hline
&&& & &&& & 1&&1& & &&& \\
&&& & &&& & &1&&1 & &&& \\
&&& & &&& & 1&&-1& & &&& \\
&&& & &&& & &1&&-1 & &&& \\
\hline
&&& & &&& & &&& & 1&&1& \\
&&& & &&& & &&& & &1&&1 \\
&&& & &&& & &&& & 1&&-1& \\
&&& & &&& & &&& & &1&&-1
\end{array}\right]
$$

$$
\mathbf{L}_3 = \left[\begin{array}{cccc|cccc|cccc|cccc}
1 &&&& C_0 &&&& &&&& &&&\\
&1&&&& C_1 && S_1 & &&&& &&&\\
&&1&&&& K_2 && &&&& &&&\\
&&&1&& S_3 && C_3 & &&&& &&&\\
\hline
1 &&&& C_4 &&&& &&&& &&&\\
&1&&&& C_5 && S_5 & &&&& &&&\\
&&1&&&& K_6 && &&&& &&&\\
&&&1&& S_7 && C_7 & &&&& &&&\\
\hline
&&&& &&&& 1 &&&& C_0 &&&\\
&&&& &&&& &1&&&& C_1 && S_1\\
&&&& &&&& &&1&&&& K_2 &\\
&&&& &&&& &&&1&& S_3 && C_3\\
\hline
&&&& &&&& 1 &&&& C_4 &&&\\
&&&& &&&& &1&&&& C_5 && S_5\\
&&&& &&&& &&1&&&& K_6 &\\
&&&& &&&& &&&1&& S_7 && C_7
\end{array}\right]
$$

$$
\mathbf{L}_4 = \left[\begin{array}{cccc|cccc|cccc|cccc}
1 &&&& &&&& C_0 &&&& &&&\\
&1&&& &&&& & C_1 &&& &&& S_1\\
&&1&& &&&& && C_2 && && S_2 &\\
&&&1 & &&&& &&& C_3 & & S_3 &&\\
\hline
&&&& 1 &&&& &&&& K_4 &&&\\
&&&& &1&&& &&& S_5 & & C_5 &&\\
&&&& &&1&& && S_6 & & && C_6 &\\
&&&& &&&1 & & S_7 && & &&& C_7\\
\hline
1 &&&& &&&& K_8 &&&& &&&\\
&1&&& &&&& & C_9 &&& &&& S_9\\
&&1&& &&&& && C_{10} && && S_{10} &\\
&&&1 & &&&& &&& C_{11} & & S_{11} &&\\
\hline
&&&& 1 &&&& &&&& K_{12} &&&\\
&&&& &1&&& &&& S_{13} & & C_{13} &&\\
&&&& &&1&& && S_{14} & & && C_{14} &\\
&&&& &&&1 & & S_{15} && & &&& C_{15}
\end{array}\right]
$$

$$
2\Phi = \begin{bmatrix}
2 &&&&&&&&&&&&&&&\\
& 1-i &&&&&&&&&&&&&& 1+i\\
&& 1-i &&&&&&&&&&&& 1+i &\\
&&& 1-i &&&&&&&&&& 1+i &&\\
&&&& 1-i &&&&&&&& 1+i &&&\\
&&&&& 1-i &&&&&& 1+i &&&&\\
&&&&&& 1-i &&&& 1+i &&&&&\\
&&&&&&& 1-i && 1+i &&&&&&\\
&&&&&&&& 2 &&&&&&&\\
&&&&&&& 1+i && 1-i &&&&&&\\
&&&&&& 1+i &&&& 1-i &&&&&\\
&&&&& 1+i &&&&&& 1-i &&&&\\
&&&& 1+i &&&&&&&& 1-i &&&\\
&&& 1+i &&&&&&&&&& 1-i &&\\
&& 1+i &&&&&&&&&&&& 1-i &\\
& 1+i &&&&&&&&&&&&&& 1-i
\end{bmatrix}
$$

In the above equations C_n and S_n are the level-dependent abbreviations for ${}^{\cos}_{\sin}(2\pi n/2^L)$, respectively, and $K_n = C_n + S_n$.

The matrix representation offers a different way of viewing the fast Hartley procedure. For example, $\mathbf{L}_4$ shows the retrograde indexing of the sine factors as elements on lines of slope 45°. $\mathbf{L}_1$ and $\mathbf{L}_2$ do not have such elements at all. One might also notice that $\mathbf{L}_1$ and $\mathbf{L}_2$ have only $2N$ nonzero elements compared with $3N$ in the limit for factors of higher level. Allowing for the time taken to access the trigonometric values, one may gain an intimate understanding of the empirical timing results of Fig. 12.2.

CONVOLUTION

In the vast majority of image processing applications, convolution is carried out between two functions of which one is symmetrical. Since two-dimensional convolution performed on an image can be reduced to one dimension by spreading the image out serially as in a television waveform, it will suffice here to speak in terms of one dimension. Under conditions where $f_1(\tau)$ has no particular symmetry and is to be convolved with $f_2(\tau)$, which is an even function, the convolution theorem for the DHT is:

$$f_1(\tau) * f_2(\tau) \quad \text{has DHT} \quad H_1(\nu)H_2(\nu).$$

In other words, the DHT of a convolution of this type is the product of the two separate Hartley transforms. Therefore, to perform convolution, we take the two DHTs, multiply them together term by term, and take the DHT again. This procedure represents an improvement over taking the two DFTs, multiplying the complex values together and inverting, since one complex multiplication $(a + jb)(c + jd) = ac - bd + j(ad + bc)$ stands for four real multiplications.

In the general case where $f_2(\tau)$ is not symmetrical, the convolution theorem has a second term. Let $H_2(\nu) = H_{2e}(\nu) + H_{2o}(\nu)$, where H_{2e} and H_{2o} are the even and odd parts of $H_2(\nu)$. [If $f_2(\tau)$ were symmetrical then the odd part of $H_{2o}(\nu)$ would be zero.] The general convolution theorem then reads

$$f_1(\tau) * f_2(\tau) = H_1(\nu)H_{2e}(\nu) + H_1(-\nu)H_{2o}(\nu).$$

This theorem is immediately deducible from the convolution theorem for the DFT.

If convolution by the transform method uses up significant time the stage of permutation may be omitted, since the resulting misordering in the transform domain affects both factors identically. If fast permutation as described in the next section is used instead of explicit bit reversal, the speedup is lessened. This refinement becomes moot for users of built-in routines for transforms or transform-based convolution who value convenience over efficiency.

When $f_1(\;)$ and $f_2(\;)$ are of comparable but not equal length, the shorter sequence may be extended with zeros in order to convolve by transform. However, with digital filtering of lengthy data $f_1(\;)$, the sequence $f_2(\;)$ that performs the filtering is much shorter; in that case direct convolution by computer is recommended (see Chapter 3).

PERMUTATION

Given a sequence of eight elements $f(i)$, where the index $i = 0, \dots 7$, permutation rearranges the elements into the order $f(0), f(4), f(2), f(6), f(1), f(5), f(3), f(7)$. Writing the original index i in binary form i_b we see that reversing i_b (third row) and reconverting to decimal form produces the permuted index j.

i	0	1	2	3	4	5	6	7
i_b	000	001	010	011	100	101	110	111
	000	100	010	110	001	101	011	111
j	0	4	2	6	1	5	3	7

How this step arises is well described by Press et al. (1990). Permutation, also called bit reversal, scrambling, or shuffling, may be implemented by repeated integer division by 2 and commonly is; but implementation by reversing binary digits turns out in retrospect to have been naïve. Permutation takes a noticeable fraction of the execution time for a fast Fourier transform as shown in the classical timing diagram of Fig. 12.6. In the example given above, half of the indices, namely 0, 2, 5, and 7, remain unchanged because their binary representations are symmetrical, an observation showing that substantial time can be saved by skipping them. Looking for further simplification I plotted scatter diagrams of j versus i and found that all the diagrams break down into two basic cells (Fig. 12.6).

The first of these calls for 4-element permutation by swapping the pair of indices, $1 \rightleftharpoons 2$, while the second involves the 8-element example tabulated above,

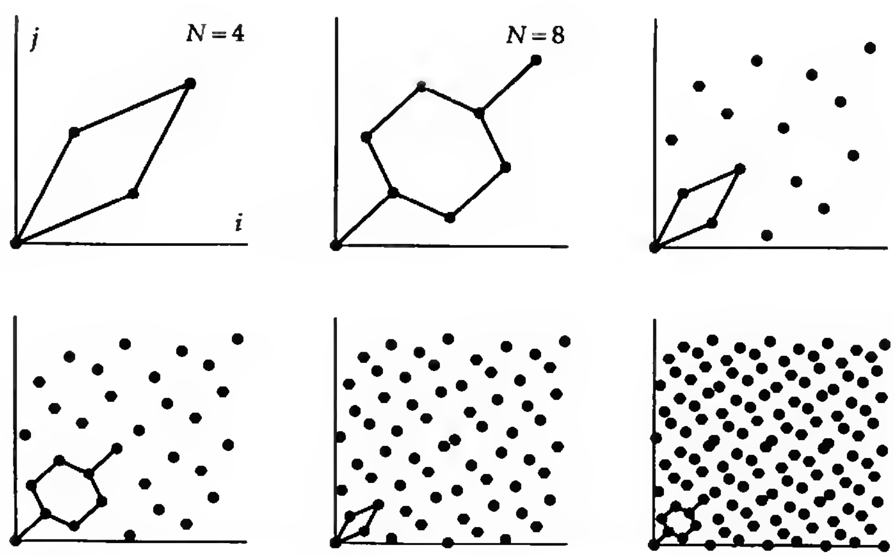


Fig. 12.6 Scatter diagrams of the permuted, or bit-reversed, index j versus the index i for $N = 4, 8, 16, \dots$ showing that each diagram is like a crystal formed by replicating cells that contain a kite ($N = 4$) or a frog ($N = 8$).

TABLE 12.4

Pseudocode for a fast Hartley subprogram in which N is a power of 4.

```
FAST HARTLEY TRANSFORM TO RADIX 4
SUB "FHTradix4" (F(),P)
N=4^P
N4=N/4
R=SQR(2)
                        Permute to radix 4
J=1
I=0
a: I=I+1
IF I>=J THEN GOTO b
T=F(J-1)
F(J-1)=F(I-1)
F(I-1)=T
b: K=N4
c: IF 3*K>=J THEN GOTO d
J=J-3*K
K=K/4
GOTO c
d: J=J+K
IF I<N-1 THEN GOTO a
                               Get DHT
                               Stage 1
FOR I=0 TO N-1 STEP 4
   T1=F(I)+F(I+1)
   T2=F(I)-F(I+1)
   T3=F(I+2)+F(I+3)
   T4=F(I+2)-F(I+3)
   F(I)=T1+T3
   F(I+1)=T1-T3
   F(I+2)=T2+T4
   F(I+3)=T2-T4
NEXT I
                           Stages 2 to P
FOR L=2 TO P
   E1=2^(L+L-3)
   E2=E1+E1
   E3=E2+E1
   E4=E3+E1
   E5=E4+E1
   E6=E5+E1
   E7=E6+E1
   E8=E7+E1
   FOR J=0 TO N-1 STEP E8
      T1=F(J)+F(J+E2)
      T2=F(J)-F(J+E2)
      T3=F(J+E4)+F(J+E6)
      T4=F(J+E4)-F(J+E6)
      F(J)=T1+T3
      F(J+E2)=T1-T3
      F(J+E4)=T2+T4
      F(J+E6)=T2-T4
      T1=F(J+E1)
      T2=F(J+E3)*R
      T3=F(J+E5)
      T4=F(J+E7)*R
      F(J+E1)=T1+T2+T3
      F(J+E3)=T1-T3+T4
      F(J+E5)=T1-T2+T3
      F(J-E7)=T1-T3-T4
      FOR K=1 TO E1-1
         L1=J+K
         L2=L1+E2
         L3=L1+E4
         L4=L1+E6
         L5=J+E2-K
         L6=L5+E2
         L7=L5+E4
         L8=L5+E6
         A1=PI*K/E4
         A2=A1+A1

         A3=A1+A2
         C1=COS(A1)
         S1=SIN(A1)
         C2=COS(A2)
         S2=SIN(A2)
         C3=COS(A3)
         S3=SIN(A3)
         T5=F(L2)*C1+F(L6)*S1
         T6=F(L3)*C2+F(L7)*S2
         T7=F(L4)*C3+F(L8)*S3
         T8=F(L6)*C1-F(L2)*S1
         T9=F(L7)*C2-F(L3)*S2
         T0=F(L8)*C3-F(L4)*S3
         T1=F(L5)-T9
         T2=F(L5)+T9
         T3=-T8-T0
         T4=T5-T7
         F(L5)=T1+T4
         F(L6)=T2+T3
         F(L7)=T1-T4
         F(L8)=T2-T3
         T1=F(L1)+T6
         T2=F(L1)-T6
         T3=T8-T0
         T4=T5+T7
         F(L1)=T1+T4
         F(L2)=T2+T3
         F(L3)=T1-T4
         F(L4)=T2-T3
      NEXT K
   NEXT J
NEXT L
SUBEND
```

which calls for swapping two pairs $1 \rightleftharpoons 4$ and $3 \rightleftharpoons 6$. These trivial swaps are then simply replicated over all the cells with no further operations than shifts. For a program see Bracewell (1986) and for other papers developing the approach see Evans (1987) and Walker (1990).

A FAST HARTLEY SUBROUTINE

As the speed of central processing units has increased year by year the benefit of small improvements in the speed of programs has diminished. However, a radix-4 algorithm that saves around 25 percent of the running time can be recommended when efficiency is important. Restrict data lengths to powers of 4 (16, 64, 256, 1024, 4096, . . .) and append zeros to data sets as needed. The following pseudocode appeared in Bracewell (1995) and is published with the permission of the American Institute of Physics.

Methods of combining radix-2 with radix-4 algorithms to gain a speed advantage are available (Pei and Wu, 1986; Bracewell, 1987). For very long data sets, parallel processing using a vector Hartley transform has been considered (Villasenor and Bracewell, 1989) and found attractive as compared with corresponding procedures for getting the FFT on a multiprocessor machine.

Short code segments elsewhere in this book are intended to be explanatory, but this subroutine is presented as a tool for spectral analysis. The pseudocode chooses expressions on the left rather than the examples on the right on the basis of which is intelligible to more readers.

```
FOR K = 0 TO N-1 STEP 4            FOR K = 1:4:N
NEXT K                        NEXT or ENDFOR or END
K < N                                   (K .LT. N)
```

When time is of the essence it seems wasteful to work around **GOTO** statements knowing that the compiler will convert the resulting indirect code back to instructions to transfer execution in just the way that the high-level **GOTO** specified. **GOTO** forms part of C, FORTRAN, and PASCAL, but not MATLAB. However the permutation segment in the pseudocode subroutine can be purged of **GOTO**s as follows, for a sequence **F( )** indexed from 1 to N (kindly supplied by John E. Baron).

```
J = 1;
I = 0;
while (1 < N - 1),
    I = I + 1;
    if (I < J),
       T = F(J);
       F(J) = F(I);
       F(I) = T;
    end;
```

```
        K = N4;
        while (3 * K < J),
           J = J - 3 * K;
           K = K / 4;
        end;
        J = J + K;
     end;
```

For fast operation, the pseudocode can be rewritten in C or in an assembler language. For easy immediate access to Hartley transforms, use an available FFT package and subtract the imaginary parts from the real parts using only half of the complex values generated.

BIBLIOGRAPHY

Ahmed, N., T. Natarajan, and K. Rao: *IEEE Trans. Comp.*, vol. C-23, pp. 90–93, 1974.

Bracewell, R. N.: "The Discrete Hartley Transform," *J. Optical Soc. Amer.*, vol. 73, pp. 1832–1835, 1983.

Bracewell, R. N.: "The Fast Hartley Transform," *Proc. IEEE.*, vol. 72, pp. 1010–1018, 1984.

Bracewell, R. N.: "The Hartley Transform," Oxford University Press, New York, 1986.

Bracewell, R. N.: "Alternative to Split-Radix Hartley Transform," *Electronics Letters*, vol. 23, pp. 1148–1149, 1987.

Bracewell, R. N.: "Physical Aspects of the Hartley Transform," *J. Atmospheric Terestrial Physics*, vol. 51, pp. 791–795, 1989.

Bracewell, R. N.: "Assessing the Hartley Transform," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. ASSP-38, pp. 2174–2176, 1990.

Bracewell, R. N.: "Affine Theorem for the Hartley Transform," *Proc. IEEE*, vol. 82, pp. 388–390, 1994.

Bracewell, R. N.: "Computing with the Hartley Transform," *Computers in Physics*, vol. 9, pp. 373–379, 1995a.

Bracewell, R. N.: "Two-Dimensional Imaging," Prentice-Hall, Englewood Cliffs, New Jersey, 1995b.

Bracewell, R. N., H. Bartelt, A. W. Lohmann, and N. Streibl: *Applied Optics*, "Optical Synthesis of Hartley Transform," vol. 24, pp. 1401–1402, 1985.

Bracewell, R. N., O. Buneman, H. Hao, and J. Villasenor: "Fast Two-dimensional Hartley Transform," *Proc. IEEE*, vol. 74, pp. 1282–1283, 1986.

Buneman, O.: "Inversion of the Helmholtz (or Laplace-Poisson) Operator for Slab Geometry," *J. Comp. Phys.*, vol. 12, pp. 124–130, 1973.

Buneman, O.: "Conversion of FFT's to Fast Hartley Transforms," *SIAM J. Sci. Statist. Comp.*, vol. 7, pp. 624–638, 1986.

Buneman, O.: "Multidimensional Hartley Transform," *Proc. IEEE*, vol. 75, p. 267, 1987.

Buneman, O.: "In-situ Bit-reversed Ordering for the Hartley Transform," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. ASSP-35, pp. 1120–1125, 1989.

Ersoy, O. K.: "A Comparative Review of Real and Complex Fourier-Related Transforms," *Proc. IEEE*, vol. 82, pp. 429–447, 1994.

Evans, D. M.: "An Improved Bit-reversal Algorithm," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. ASSP-35, pp. 1120–1125, 1987.

Guo, H., G. A. Sutton, and C. S. Burrus: "The Quick Fourier Transform," *IEEE Trans. SP*, vol. 46, pp. 338–341, 1998.

Hao, H., and R. N. Bracewell: "A Three-dimensional DFT Using the Fast Hartley Transform," *Proc. IEEE*. vol. 75, pp. 264–266, 1987.

Hartley, R. V. L.: "A More Symmetrical Fourier Analysis Applied to Transmission Problems," *Proc. Inst. Radio Engrs.*, vol. 30, pp. 144–150, 1942. Introduced the cas function kernel, emphasizing the property of real spectral representation of real data and noting the failure of reciprocity of the complex Fourier transform kernel.

Hou, H. S.: "The Fast Hartley Transform Algorithm," *IEEE Trans. Comp.*, vol. C-36, pp. 147–156, 1987.

Meher, P. K., J. K. Satapathy, and G. Panda: *Electronics Letters*, vol. 28, pp. 1077–1078, 1992.

Millane, R. P.: "Analytic Properties of the Hartley Transform and Their Implications," *Proc. IEEE*, vol. 82, pp. 413–428, 1994.

Olnejniczak, K. J.: "The Hartley Transform," in E. Poularikas ed. "The Handbook of Transforms," Chemical Rubber Company, Boca Raton, Florida, 1998.

Olnejniczak, K. J., and G. T. Heydt: "Scanning the Special Section on the Hartley Transform," *Proc. IEEE*, vol. 82, pp. 372–380, 1994. Contains extensive bibliography.

Pei, S-C, and J-L Wu: "Split-Radix Fast Hartley Transform," *Electronics Letters*, vol. 22, pp. 26–27, 1986.

Pennebaker, W., and J. L. Mitchell: "JPEG: Still Image Data Compression Standard," Van Nostrand Reinhold, New York, 1993.

Perkins, M. G.: "A Separable Hartley-like Transform in Two or More Dimensions," *Proc. IEEE*, vol. 75, pp. 1127–1129, 1987.

Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling: "Numerical Recipes, the Art of Scientific Programming," Cambridge University Press, Cambridge, UK, 1990. See also successor volumes devoted to programs and examples in BASIC, C, FORTRAN 77, FORTRAN 90, and PASCAL.

Rao, K. R., and P. Yip: "Discrete Cosine Transform: Algorithms, Advantages, Applications," Academic Press, Boston, 1990.

Trans. IEEE. Special issue on the fast Fourier transform with an extensive bibliography. *IEEE Trans. Audio and Electroacoustics*, vol. AU-17, no. 2, June 1969.

Villasenor, J. D.: "Two-dimensional Optical Hartley Transforms in the Presence of Errors," *Applied Optics*, vol. 28, pp. 2671–2676, 1989.

Villasenor, J. D.: "Optical Hartley Transform," *Proc. IEEE*, vol. 82, pp. 391–399, 1994.

Villasenor, J. D., and R. N. Bracewell: "Optical Phase Obtained by Analogue Hartley Transformation," *Nature*, vol. 330, pp. 735–737, 1987.

Villasenor, J. D., and R. N. Bracewell: "Lensless Microwave Imaging Using the Hartley Transform," *Nature*, vol. 335, pp. 617–619, 1988.

Villasenor, J. D., and R. N. Bracewell: "Vector Hartley Transform," *Electronics Letters*, vol. 25, pp. 1110–1111, 1989.

Walker, J. S.: "A New Bit-reversal Algorithm," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. ASSP-38, pp. 1472–1474, 1990.

Wang, Z.: "Fast Algorithms for the Discrete W Transform and for the Discrete Fourier Transform," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. ASSP-32, pp. 803–816, 1984.

Yang, D.: "New Fast Algorithm to Compute the Two-dimensional Discrete Hartley Transform," *Electronics Letters*, vol. 25, pp. 1705–1706, 1989.

PROBLEMS

1. Obtain the DHTs of the following sequences: {1 2}, {1 2 3 4}, {1 2 3 4 5 6 7 8}.

2. (*a*) From the previous results guess a rule about the occurrence of a zero element in the DHT of $\{1\ 2\ 3\ \ldots\ 2^P\}$.
   (*b*) Is your rule true for $P = 4$?
   (*c*) Prove algebraically that the DHT of $\{1\ 2\ 3\ \ldots\ 2^P\}$ always has one term that is zero.

3. Obtain the DHTs of the following sequences: {2 1}, {4 3 2 1}, {8 7 6 5 4 3 2 1}.

4. (*a*) Obtain the DFT of {8 7 6 5 4 3 2 1} from the DHT. (b) Deduce the power spectrum of {8 7 6 5 4 3 2 1} directly from the DHT.

5. Obtain the DHTs of {2 3 4 5 6 7 8 1} and {3 4 5 6 7 8 1 2}.

6. (*a*) Deduce the DFT of {3 4 5 6 7 8 1 2} from the DHT.
   (*b*) Deduce the power spectrum of {3 4 5 6 7 8 1 2} directly from the DHT.

7. Obtain the DHTs of {7 6 5 4 3 2 1 8} and {6 5 4 3 2 1 8 7}.

8. Obtain the DHT of {1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4}.

9. (*a*) Obtain the DHTs of {1 2 1 2 1 2 1 2} and {1 1 2 1 1 2 1 1}.
   (*b*) Explain why the peaks fall where they do.

10. (*a*) Obtain the DHT of $\{-2\ \ 2\ \ -2\ \ 2\ \ -2\ \ 2\ \ -2\ \ 2\}$ and explain the structure of the DHT.
    (*b*) Obtain the power spectrum and explain the structure. ▷

11. **New use for cas function.** Make a polar coordinate plot of $r = [\mathrm{cas}(\theta \bmod \tfrac{1}{2}\pi)]^{-1}$, for $0 \leqslant \theta \leqslant 2\pi$. ▷

12. **Cyclic transforms.** Show that application of the Fourier transform four times in succession recovers the original function, i.e., that $\mathcal{F}\mathcal{F}\mathcal{F}\mathcal{F}f(x) = f(x)$, and that for the Hartley transform $\mathcal{H}\mathcal{H}f(x) = f(x)$. ▷

13. **Discrete cosine transform.** Taking $N = 8$, graph the basis functions versus τ, for $\nu = 0, 1, 2, 3$ and for $\nu = 7$, for each of the discrete transforms DCT1 and DCT2. Select a pair of basis functions and check whether the sum of the products is zero. ▷

14. **Inserted zeros.** Obtain the DFTs for the following sequences and point out any noticeable interrelationships.

    {2718 2818 0000 0000}, {0000 2718 2818 0000}, {2000 0000 0818 2817}.

15. **Integer data.** Under what condition will a sequence of eight integers have a DHT consisting of integers only?

16. **Binomial sequences.** Comment on the DHTs of the following sequences.

 {1 7 21 35 35 21 7 1}, {35 35 21 7 1 1 7 21}, {1 8 28 56 70 56 28 8 1}, {70 56 28 8 1 1 8 28 56}.

17. **Random binary data.** A sequence consists of N randomly chosen values 1 or -1. (*a*) What are the expected values of $H(0)$ and $H(N/2)$? (*b*) What would you expect for the standard deviation of $H(0)$ if values were computed for a substantial number of such sequences?

18. **Accuracy of discrete transform.** Gain some numerical intuition by comparing the DHT of {8 7 6 5 4 3 2 1 0 0 0 0 0 0 0 0 0 1 2 3 4 5 6 7} with 24 equispaced values of $(8/3)\text{sinc}^2(\nu/3)$.

19. **Sum of two angles.** Show that

$$\text{cas}(A + B) = \text{cas}\, A \,\text{cas}\, B + \text{cas}'\, A \sin B.$$

20. **Derivative theorem.** If $V(t)$ has Hartley transform $H(f)$, what is the Hartley transform of $V'(t)$? ▷

21. **Hartley intensity.** The Hartley *intensity* $[H(u,\nu)]^2$ of a shifted impulse $\delta(x - a, y - b)$ is seen to consist of parallel fringes running from WNW to SSE. The fringe crest that passes nearest to the origin does so in the first quadrant and the fringe period is d. What is the exact location (a, b) of the original impulse? ▷

22. **Complex Hartley plane.** Show that the complex plane whose real axis is $H(\nu)$ and imaginary axis $H(-\nu)$ is related to the complex Fourier phase by a $-45°$ rotation and magnification by $\sqrt{2}$. In other words

$$H(\nu) + iH(-\nu) = (F_{real} + iF_{imag}) \times \sqrt{2}e^{i\pi/4}. \triangleright$$

23. **Fourier phase.** Show that the Fourier phase $\phi = \arctan(F_{imag}/F_{real})$ is deducible directly from the Hartley transform without reference to the complex Fourier transform at all by $\phi = \arctan[H(-\nu)/H(\nu)] - \pi/4$. ▷

24. **Affine theorem.** The affine transformation in two dimensions (Bracewell, 1993) states that if $f(x,y) \overset{2}{\supset} H(u,\nu)$ then

$$f(ax + by + c, dx + ey + f) \overset{2}{\supset} \frac{H(\alpha,\beta)\cos\Theta - H(-\alpha,-\beta)\sin\Theta}{|ae - bd|},$$

 where

$$\alpha = \frac{eu - dv}{ae - bd}, \quad \beta = \frac{-bu + av}{ae - bd}, \text{ and } \quad \Theta = \frac{2\pi[(ec - bf)/u + (af - cd)/v]}{ae - bd}.$$

 Many of the Hartley transform theorems are special cases. Show that the rotation theorem is a special case of the affine theorem. ▷

25. **Convolution by DCT.** Hubert Smith, MD, and Dy Jones, MS, work as programmers in a small medical instrumentation company. Smith says, "I want to convolve using the DCT, but I am too busy to do the algebra. I know that the DCT of a convolution is

K times the product of the DCTs and I know that $\{4\,3\,2\,1\}$ has DCT $\{5.0000\;\; 2.2304\;\; 0\;\; 0.1585\}$ and that $\{2\,3\,5\,7\}$ has DCT $\{8.5000\;\; -3.8076\;\; 0.5000\;\; -0.0464\}$. Do you think this will give me K?"

Jones says, "Shouldn't you avoid cyclic overlap by starting with $\{1\,2\,3\,4\,0\,0\,0\}$ and $\{2\,3\,5\,7\,0\,0\,0\}$?" What value of K did these workers arrive at, and what advice do you have for them?

26. Checking the DCT numerically. Discrete Hartley transforms can be checked numerically by seeing that the initial element $H(0)$ is the mean, or d.c., value of the data. Conversely, the inverse transformation can be checked by comparing $f(0)$ with the sum of the DHT values. The same is true of the DFT, taking into account that the values are complex. When it comes to checking the discrete transform we note that the DCT2 of $\{1\,2\,3\,4\}$ is $\{1.77\;\; -0.79\;\; 0\;\; -0.06\}$ and that ${}^{\mathrm{II}}F(0)$ is neither the sum nor the mean of the values of $f(\tau)$. What is the expression that checks ${}^{\mathrm{II}}F(0)$? Conversely, what property of ${}^{\mathrm{II}}F(\nu)$ can be used to check $f(0)$?

Now noting that the DCT1 of $f(\tau) = \{1\,2\,3\,4\}$ is ${}^{\mathrm{I}}F_c(\nu) = \{4.03\;\; 2.14\;\; 0.85\;\; -0.65\}$, what are the expressions for checking the leading elements of $f(\tau)$ and ${}^{\mathrm{II}}F_c(\nu)$? ▷

13

Relatives of the Fourier Transform

Many of the linear transforms in common use have a direct connection with either the Fourier or the Laplace transform. The closest relationship is with the generalizations of the Fourier transform to two or more dimensions, and with the Hankel transforms of the zero and higher orders, into which the multidimensional Fourier transforms degenerate under circumstances of symmetry. The Mellin transform is illuminated by previous study of the Laplace transform and is the tool by which the fundamental theory of Fourier kernels is constructed. Particularly impressive is the simplification of the Hilbert transform when it is studied by the Fourier transform, and finally there is the intimate relationship whereby the Abel, Fourier, and Hankel transformations, applied in succession, regenerate the original function.

All these transforms are tabulated in two volumes by Erdélyi et al. (1954). The Radon transform (Deans, 1983) may be viewed as a generalization of the Abel transform where circular symmetry is dropped. The Abel-Fourier-Hankel cycle then generalizes to the projection-slice theorem, an important thinking tool for reconstruction from projections, or inversion of the Radon transform.

THE TWO-DIMENSIONAL FOURIER TRANSFORM

The variable x may stand for some physical quantity such as time or frequency, which is essentially one-dimensional, or it may be the coordinate in a one-dimensional physical system such as a stretched string or an electrical transmission line. However, in cases which are two dimensional—stretched membranes, antennas and arrays of antennas, lenses and diffraction gratings, pictures on television screens, and so on—more general formulas apply.

A two-dimensional function $f(x,y)$ has a two-dimensional transform $F(u,v)$, and between the two the following relations exist:

$$F(u,v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)e^{-i2\pi(ux+vy)}\,dx\,dy$$

$$f(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(u,v)e^{i2\pi(ux+vy)}\,du\,dv.$$

These equations describe an analysis of the two-dimensional function $f(x,y)$ into components of the form $\exp[i2\pi(ux + vy)]$. Since any such component can be split into cosine and sine parts, we may begin by considering a cosine component $\cos[2\pi(ux + vy)]$.

As an example of a two-dimensional function consider the height of the ground at the geographical point (x,y), for example, over the area occupied by the mountain which is conventionally represented in Fig. 13.1 by contours of constant height. The function $\cos[2\pi(ux + vy)]$ represents a cosinusoidally corrugated land surface whose contours of constant height coincide with lines whose equation is

$$ux + vy = \text{const.}$$

The corrugations face in a direction that makes an angle arctan (v/u) with the x axis and their wavelength is $(u^2 + v^2)^{-\frac{1}{2}}$. If a section is made through the corrugations, in the x direction, it will undulate with a frequency of u cycles per unit of x. Similarly, v may be interpreted as the number of cycles per unit of y, in the y direction.

In Fig. 13.1, a prominent Fourier component of the mountain is shown. In the transform domain the complex component is characterized in wavelength and orientation by the point (u,v) in the uv plane and its amplitude by $F(u,v)$. The interpretation of u and v as spatial frequencies is emphasized by dimensioning u^{-1} and v^{-1}, the wavelengths of sections taken in the x and y directions, respectively (see Fig. 13.1). The second of the Fourier relations quoted above asserts that a summation of corrugations of appropriate wavelengths and orientations, taken with suitable amplitudes, can reproduce the original mountain. The sinusoidal components, which must also be included, allow for the possibility that the corrugations may have to be slid into appropriate spatial phases.

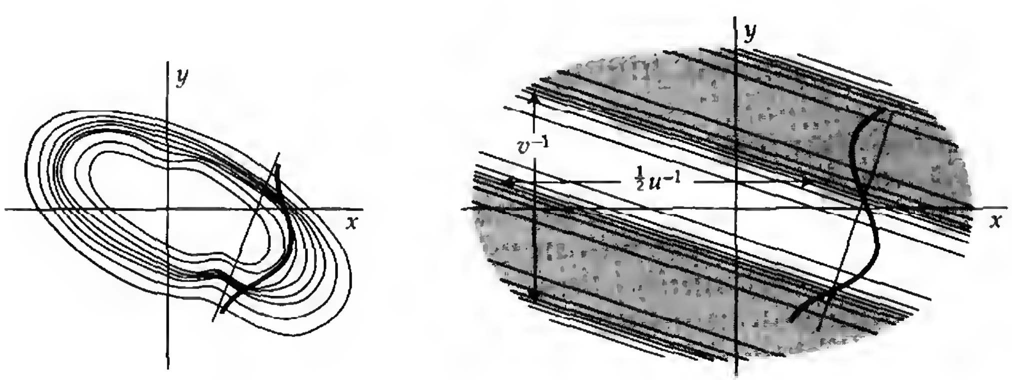


Fig. 13.1 A mountain (left) and a prominent Fourier component thereof (right).

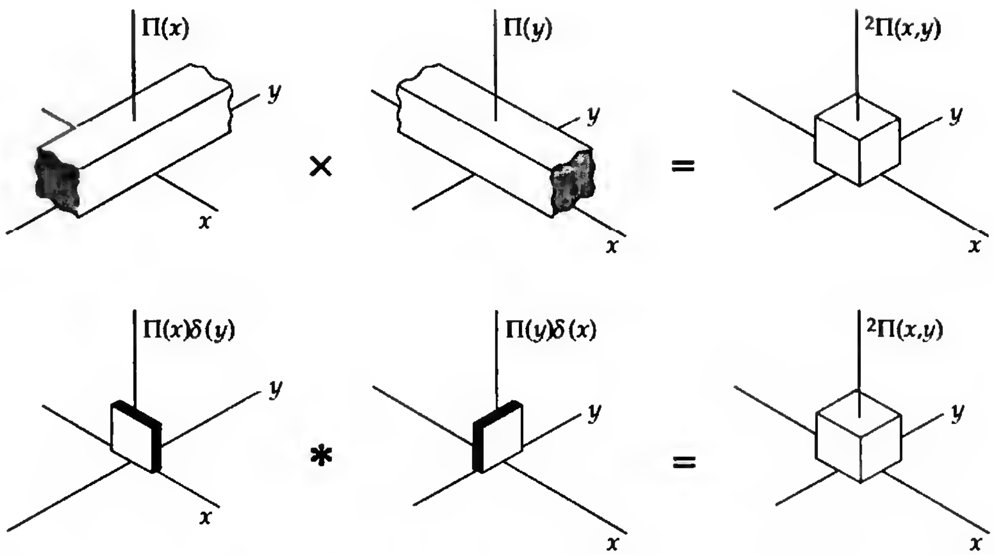


Fig. 13.2 **Expressing a two-dimensional function as a product and as a convolution.**

TWO-DIMENSIONAL CONVOLUTION

The convolution integral of two two-dimensional functions $f(x,y)$ and $g(x,y)$ is defined by

$$f ** g = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x',y')g(x - x', y - y')\, dx'\, dy'.$$

Thus one of the functions is rotated half a turn about the origin by reversing the sign of both x and y, displaced, and multiplied with the other function, and the product is then integrated to obtain the value of the convolution integral for that particular displacement.

The two-dimensional autocorrelation function is formed in the same way save that the sign reversal is omitted; thus

$$f \star\star g = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x',y')g(x + x', y + y')\, dx'\, dy'.$$

It is often convenient to be able to perceive ways in which a given function can be expressed as a convolution. For example, the two-dimensional function

$$^2\Pi(x,y) = \begin{cases} 1 & |x| \text{ and } |y| < \frac{1}{2} \\ 0 & \text{elsewhere} \end{cases}$$

may be expressed as a product or as a convolution:

$$^2\Pi(x,y) = \Pi(x)\Pi(y) = [\Pi(x)\,\delta(y)] * [\Pi(y)\,\delta(x)].$$

These two possibilities are illustrated in Fig. 13.2.

TABLE 13.1
Theorems for the two-dimensional Fourier transform

Theorem	$f(x,y)$	$F(u,v)$
Similarity	$f(ax,by)$	$\dfrac{1}{\lvert ab\rvert} F\left(\dfrac{u}{a}, \dfrac{v}{b}\right)$
Addition	$f(x,y) + g(x,y)$	$F(u,v) + G(u,v)$
Shift	$f(x-a, y-b)$	$e^{-2\pi i(au+bv)}F(u,v)$
Modulation	$f(x,y) \cos \omega x$	$\frac{1}{2}F\left(u + \dfrac{\omega}{2\pi}, v\right) + \frac{1}{2}F\left(u - \dfrac{\omega}{2\pi}, v\right)$
Convolution	$f(x,y) * g(x,y)$	$F(u,v)G(u,v)$
Autocorrelation	$f(x,y) * f^*(-x,-y)$	$\lvert F(u,v)\rvert^2$
Rayleigh	$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \lvert f(x,y)\rvert^2\, dx\, dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \lvert F(u,v)\rvert^2\, du\, dv$	
Power	$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)g^*(x,y)\, dx\, dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(u,v)G^*(u,v)\, du\, dv$	
Parseval	$\int_{-\frac{1}{2}}^{\frac{1}{2}}\int_{-\frac{1}{2}}^{\frac{1}{2}} \lvert f(x,y)\rvert^2 = \sum\sum {a_{mn}}^2,$ where $F(u,v) = \sum\sum a_{mn}[{}^2\delta(u-m, v-n)]$	
Differentiation	$\left(\dfrac{\partial}{\partial x}\right)^m\left(\dfrac{\partial}{\partial y}\right)^n f(x,y)$	$(2\pi iu)^m(2\pi iv)^nF(u,v)$
	$\dfrac{\partial}{\partial x} f(x,y) = f'_x(x,y)$	$2\pi iuF(u,v)$
	$\dfrac{\partial}{\partial y} f(x,y) = f'_y(x,y)$	$2\pi ivF(u,v)$
	$\dfrac{\partial^2}{\partial x^2} f(x,y) = f''_{xx}(x,y)$	$-4\pi^2u^2F(u,v)$
	$\dfrac{\partial^2}{\partial y^2} f(x,y) = f''_{yy}(x,y)$	$-4\pi^2v^2F(u,v)$
	$\dfrac{\partial^2}{\partial x\, \partial y} f(x,y) = f''_{xy}(x,y)$	$-4\pi^2uvF(u,v)$
	$\left(\dfrac{\partial^2}{\partial x^2} + \dfrac{\partial^2}{\partial y^2}\right) f(x,y)$	$-4\pi^2(u^2 + v^2)F(u,v)$

The theorems pertaining to the one-dimensional transform generalize readily, as shown briefly in Table 13.1.

Some theorems that come into existence in two dimensions are as follows.

Rotation theorem. If $f(x,y)$ is rotated on the (x,y)-plane then its transform is rotated on the (u,v)-plane through the same angle and in the same sense:

$f(x \cos \theta - y \sin \theta, x \sin \theta + y \cos \theta)$ has 2-DFT $F(u \cos \theta - v \sin \theta, u \sin \theta + v \cos \theta)$.

TABLE 13.1
Theorems for the two-dimensional Fourier transform ***(cont.)***

Theorem	$f(x,y)$	$F(u,v)$
Definite integral	$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = F(0,0)$	
First moments	$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf(x,y)\,dx\,dy = \frac{1}{-2\pi i}F'_u(0,0)$ $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x\cos\theta + y\sin\theta)f(x,y)\,dx\,dy$ $= \frac{1}{-2\pi i}[\cos\theta\, F'_u(0,0) + \sin\theta\, F'_v(0,0)]$	
Center of gravity	$\langle x\rangle = \frac{F'_u(0,0)}{-2\pi i F(0,0)}$ $\quad\langle y\rangle = \frac{F'_v(0,0)}{-2\pi i F(0,0)}$	
Second moments	$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^2f(x,y)\,dx\,dy = \frac{F''_{uu}(0,0)}{-4\pi^2}$ $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)\,dx\,dy = \frac{F''_{uv}(0,0)}{-4\pi^2}$ $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x^2+y^2)f(x,y)\,dx\,dy = -\frac{1}{4\pi^2}[F''_{uu}(0,0) + F''_{vv}(0,0)]$	
Equivalent width	$\frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy}{f(0,0)} = \frac{F(0,0)}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(u,v)\,du\,dv}$	
Finite differences†	$\Delta_x f(x,y)$ $\Delta^2_{xy} f(x,y)$ $\Delta^2_{xx} f(x,y)$	$i2\sin\pi u\, F(u,v)$ $-4\sin\pi u\sin\pi v\, F(u,v)$ $-4(\sin\pi u)^2F(u,v)$
Running means	$\left[\Pi\left(\frac{x}{a}\right)\Pi\left(\frac{y}{b}\right)\right] * f(x,y)$	$ab\,\mathrm{sinc}\,au\,\mathrm{sinc}\,bv\,F(u,v)$
Separable product	$f(x)g(y)$	$F(u)G(v)$

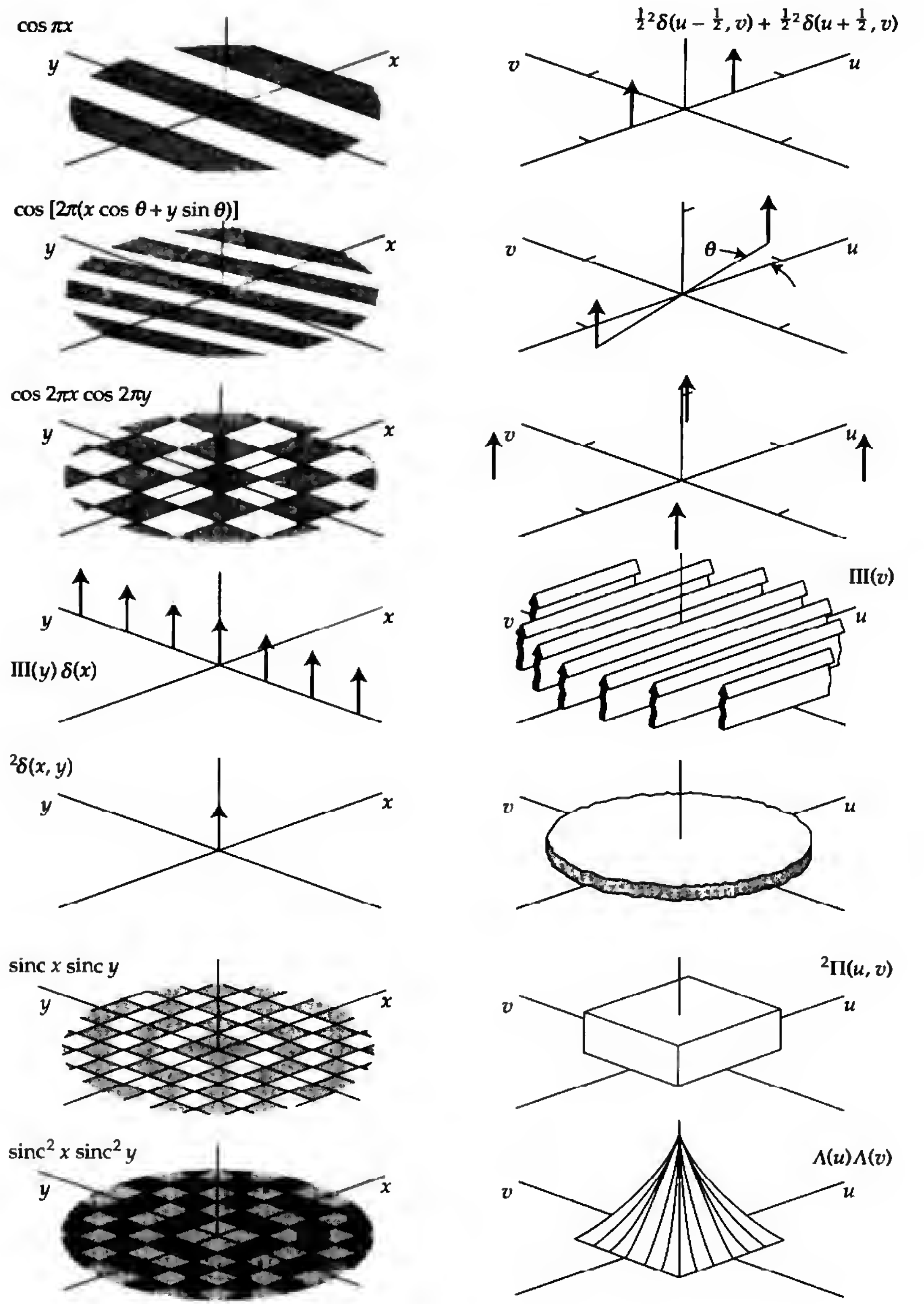
†The finite differences in the table are defined as follows:
$\Delta_x f(x,y) = f(x+\frac{1}{2}, y) - f(x-\frac{1}{2}, y)$
$\Delta^2_{xy} f(x,y) = f(x+\frac{1}{2}, y+\frac{1}{2}) - f(x-\frac{1}{2}, y+\frac{1}{2}) - f(x+\frac{1}{2}, y-\frac{1}{2}) + f(x-\frac{1}{2}, y-\frac{1}{2})$
$\Delta^2_{xx} f(x,y) = f(x+1, y) - 2f(x,y) + f(x-1, y).$

Simple shear theorem. If $f(x,y)$ is subjected to shear then its transform is sheared to the same degree in the perpendicular direction:

$$f(x+by, y) \text{ has 2-D FT } F(u, v-bu).$$

Affine theorem. A function $f(x,y)$ subjected to an affine transformation of its coordinate plane, becomes (Bracewell et al., 1993; Bracewell, 1994) $f(ax+by+c, dx+ey+f)$ and then has 2-D FT $|ae-bd|^{-1}\exp\{i2\pi(ae-bd)^{-1}[(ec-bf)u+(af-cd)v]\}F[(eu-dv)/(ae-bd), (-bu+av)/(ae-bd)]$.

Figure 13.3 illustrates a number of two-dimensional Fourier transforms.



$\cos \pi x$
$\frac{1}{2}{}^2\delta(u-\frac{1}{2}, v) + \frac{1}{2}{}^2\delta(u+\frac{1}{2}, v)$
$\cos [2\pi(x \cos \theta + y \sin \theta)]$
θ
$\cos 2\pi x \cos 2\pi y$
$\text{III}(y)\,\delta(x)$
$\text{III}(v)$
${}^2\delta(x, y)$
$\operatorname{sinc} x \operatorname{sinc} y$
${}^2\Pi(u, v)$
$\operatorname{sinc}^2 x \operatorname{sinc}^2 y$
$\Lambda(u)\Lambda(v)$
x
y
u
v

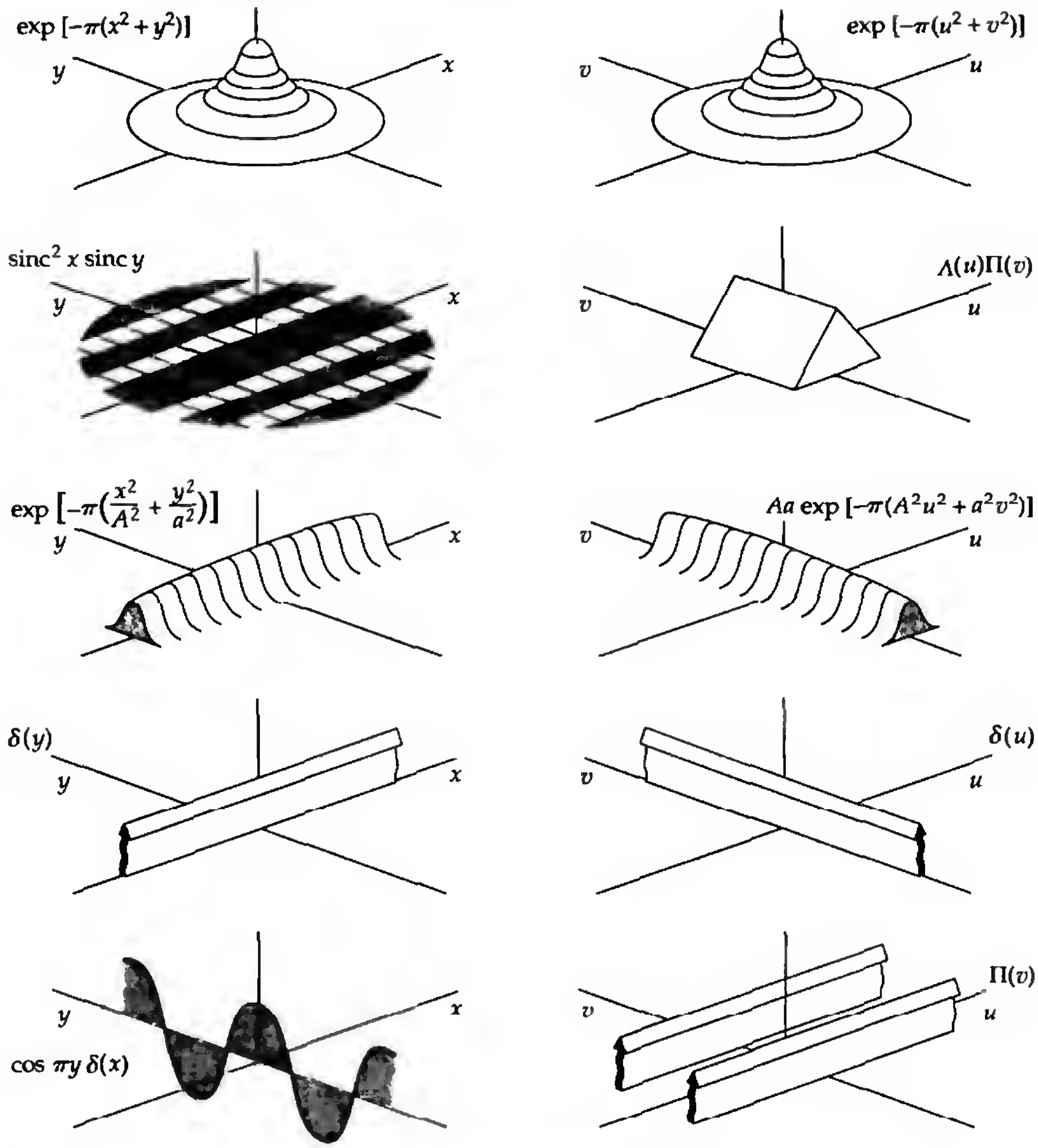


Fig. 13.3 **Some two-dimensional Fourier transforms.**

THE HANKEL TRANSFORM

Two-dimensional systems may often show circular symmetry; for example, optical systems are often constructed from components that, in themselves, are circularly symmetrical. Then again, waves spreading out in two dimensions from a source of energy exhibit symmetry for natural reasons. It may be expected that in these cases a simplification will result, for one radial variable will suffice in place

of the two independent variables x and y. The appropriate expression of such problems is in terms of the Hankel transform, a one-dimensional transform with Bessel function kernel.

When circular symmetry exists, that is, when

$$f(x,y) = \mathbf{f}(r),$$

where

$$r^2 = x^2 + y^2,$$

then $F(u,v)$ proves also to be circularly symmetrical; that is,

$$F(u,v) = \mathbf{F}(q),$$

where

$$q^2 = u^2 + v^2.$$

To show this, change the transform formula to polar coordinates and integrate over the angular variable.[1] Then the relations between the two one-dimensional functions $\mathbf{f}(r)$ and $\mathbf{F}(q)$ are

$$\mathbf{F}(q) = 2\pi \int_0^\infty \mathbf{f}(r) J_0(2\pi qr) r\, dr$$

and

$$\mathbf{f}(r) = 2\pi \int_0^\infty \mathbf{F}(q) J_0(2\pi qr) q\, dq.$$

We refer to $\mathbf{F}(q)$ as the Hankel transform (of zero order) of $\mathbf{f}(r)$ and note that the transformation is strictly reciprocal, as was the case when the kernels were cos and sin. The kernel J_0, together with cos, sin, and others, is referred to as a Fourier kernel in the broad sense of a kernel associated with a reciprocal transform.

The factors 2π in the above formulas may be canceled by suitable redefinition of the variables, but their retention follows logically from the form adopted for Fourier transforms. In physical situations the 2π in parentheses will be found to result from the measurement of q in whole cycles per unit of r. The 2π before the integral sign comes from the element of area $2\pi r\, dr$.

[1] That is,

$$\begin{aligned}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)e^{-i2\pi(xu+yv)}\,dx\,dy &= \int_0^\infty \int_0^{2\pi} \mathbf{f}(r)e^{-i2\pi qr\cos(\theta-\phi)} r\,dr\,d\theta \\ &= \int_0^\infty \mathbf{f}(r)\left[\int_0^{2\pi} e^{-i2\pi qr\cos\theta}\,d\theta\right] r\,dr \\ &= 2\pi \int_0^\infty \mathbf{f}(r)J_0(2\pi qr) r\,dr\end{aligned}$$

where $x + iy = re^{i\theta}$, $u + iv = qe^{i\phi}$ and we have used the relation

$$J_0(z) = \frac{1}{2\pi}\int_0^{2\pi} e^{-iz\cos\beta}\,d\beta.$$

A number of zero-order Hankel transforms are shown as two-dimensional Fourier transforms in Fig. 13.4. Table 13.2 lists various Hankel transforms for reference.

Many of the theorems for the two-dimensional Fourier transform can be restated in terms of the Hankel transform. The names of the corresponding Fourier theorems are listed in Table 13.3 to allow comparison.

Numerical evaluation of the Hankel transform is straightforward if the J_0 Bessel function is available. An alternative method is to obtain the Abel transform $f_A(x)$ for nonnegative x, turn it into an even function of double length by supplying values for negative x, and call a fast Fourier transform.

For aspects of the Hankel and two-dimensional Fourier transforms related to imaging see Bracewell (1995).

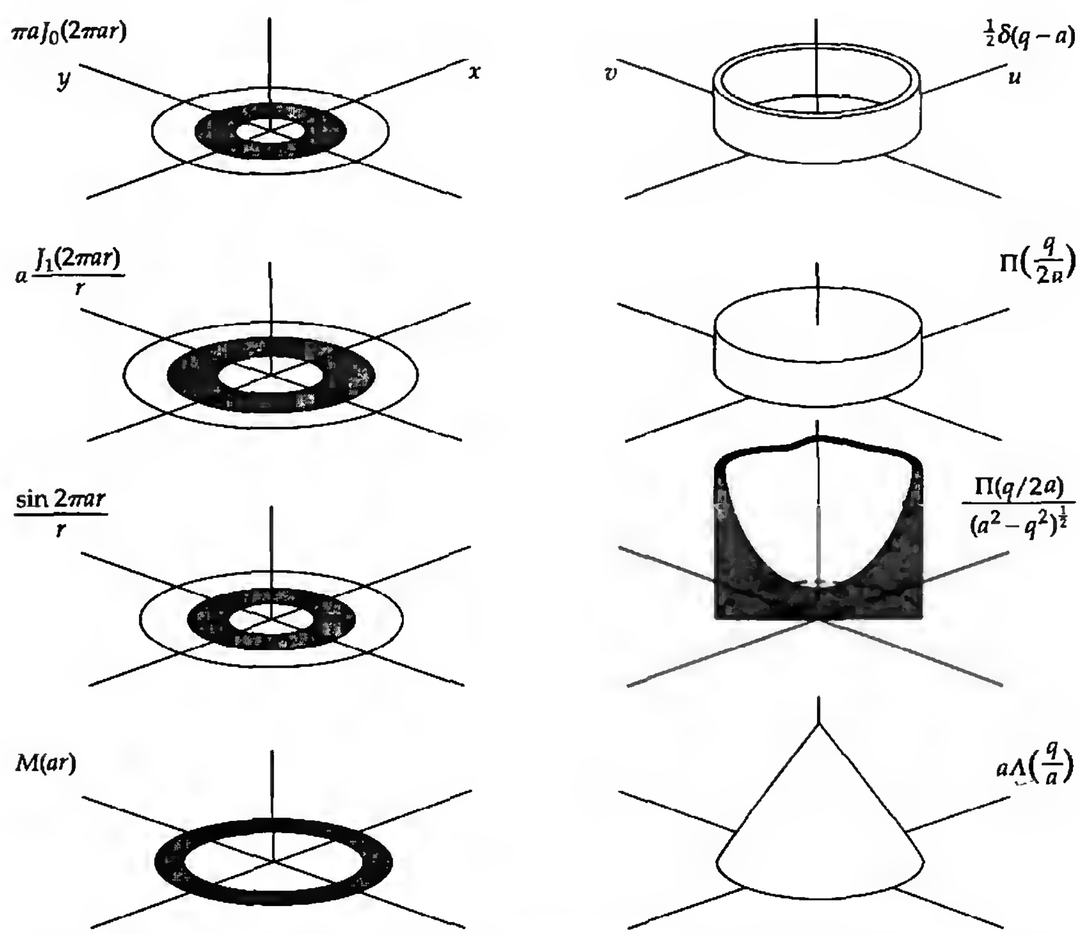


Fig. 13.4 **Some zero-order Hankel transforms shown as two-dimensional Fourier transforms.**

TABLE 13.2
Some Hankel transforms

$f(r)$	$F(q)$
$\Pi\left(\frac{r}{2a}\right)$	$\frac{aJ_1(2\pi aq)}{q}$
$\frac{\sin 2\pi ar}{r}$	$\frac{\Pi(q/2a)}{(a^2-q^2)^{\frac{1}{2}}}$
$\frac{1}{2}\,\delta(r-a)$	$\pi aJ_0(2\pi aq)$
$M(ar)$†	$a^{-2}\Lambda\left(\frac{q}{a}\right)$
$e^{-\pi r^2}$	$e^{-\pi q^2}$
$\frac{1}{(a^2+r^2)^{\frac{1}{2}}}$	$\frac{e^{-2\pi aq}}{q}$
$\frac{1}{(a^2+r^2)^{\frac{3}{2}}}$	$\frac{2\pi e^{-2\pi aq}}{a}$
$\frac{1}{a^2+r^2}$	$2\pi K_0(2\pi aq)$
$\frac{2a^2}{(a^2+r^2)^2}$	$4\pi^2 aqK_1(2\pi aq)$
$\frac{4a^4}{(a^2+r^2)^3}$	$4\pi^2 aqK_1(2\pi aq) + 4\pi^3 a^2q^2K_0(2\pi aq)$
$(a^2-r^2)\,\Pi\left(\frac{r}{2a}\right)$	$\frac{a^2J_2(2\pi aq)}{\pi q^2}$
$\frac{1}{r}$	$\frac{1}{q}$
e^{-ar}	$\frac{2\pi a}{(4\pi^2q^2+a^2)^{\frac{3}{2}}}$
$\frac{e^{-ar}}{r}$	$\frac{2\pi}{(4\pi^2q^2+a^2)^{\frac{1}{2}}}$
$\frac{\delta(r)}{\pi r} = {}^2\delta(x,y)$	1
$\frac{\Pi\left(\frac{r}{2a}\right)^{**2}}{2a^2} = \left[\cos^{-1}\frac{r}{2a} - \frac{r}{2a}\left(1-\frac{r^2}{4a^2}\right)^{\frac{1}{2}}\right]\Pi\left(\frac{r}{4a}\right)$	$\frac{[J_1(2\pi aq)]^2}{2q^2}$
$-4\pi^2r^2f(r)$	$\left(\frac{d^2F}{dq^2} + \frac{1}{q}\frac{dF}{dq}\right) = \nabla^2F$
$r^2e^{-\pi r^2}$	$\left(\frac{1}{\pi} - q^2\right)e^{\ \pi q^2}$
$J_0^2(2\pi ar)$	$\pi^{-2}q^{-1}(\pi^{-2}-q^2)^{-1/2}]\Pi(\pi q/2)$
jinc ar	$a^{-2}\Pi(q/a)$
jinc^2 ar	$\frac{1}{2}[\cos^{-1}q - q(1-q^2)^{1/2}\Pi(q/2)]$
sinc^2 ar	$(\pi a^2)^{-1}\cosh^{-1}(a/q)$
$r^{-n}J_n(r)$	$2\pi(1-4\pi^2q^2)^{n-1}\Pi(\pi q)/2^{n-1}(n-1)!$
e^{ir^2}	$i\pi e^{-i\pi^2q^2}$

†In this table $M(x) = 2\pi\left[x^{-3}\int_0^x J_0(x)\,dx - x^{-2}J_0(x)\right]$.

TABLE 13.3
Theorems for the Hankel transform

Theorem	f(r)	F(q)
Similarity	$\mathbf{f}(ar)$	$a^{-2}\mathbf{F}\left(\frac{q}{a}\right)$
Addition	$\mathbf{f}(r) + \mathbf{g}(r)$	$\mathbf{F}(q) + \mathbf{G}(q)$
Shift	Shift of origin destroys circular symmetry	
Convolution	$\int_0^\infty \int_0^{2\pi} \mathbf{f}(r')\mathbf{g}(R)r'\,dr'\,d\theta$ $(R^2 = r^2 + r'^2 - 2rr' \cos\theta)$	$\mathbf{F}(q)\mathbf{G}(q)$
Rayleigh	$\int_0^\infty \lvert\mathbf{f}(r)\rvert^2 r\,dr = \int_0^\infty \lvert\mathbf{F}(q)\rvert^2 q\,dq$	
Power	$\int_0^\infty \mathbf{f}(r)\mathbf{g}^*(r)r\,dr = \int_0^\infty \mathbf{F}(q)\mathbf{G}^*(q)q\,dq$	
Differentiation	Exercise for student	
Definite integral	$2\pi \int_0^\infty \mathbf{f}(r)r\,dr = \mathbf{F}(0)$	
Second moment	$2\pi \int_0^\infty r^2\mathbf{f}(r)r\,dr = \dfrac{\mathbf{F}''(0)}{-2\pi^2}$	
Equivalent width	$\dfrac{2\pi \int_0^\infty \mathbf{f}(r)r\,dr}{\mathbf{f}(0)} = \dfrac{\mathbf{F}(0)}{2\pi \int_0^\infty \mathbf{F}(q)q\,dq}$	

FOURIER KERNELS

Let two functions f and g be related through the following integral equation whose kernel is k:

$$g(s) = \int_0^\infty f(x)k(s, x)\,dx,$$

and let the kernel be such that a reciprocal relationship also holds; that is,

$$f(x) = \int_0^\infty g(s)k(s,x)\,ds.$$

We know that 2 cos $2\pi ax$, 2 sin $2\pi ax$, and cas $2\pi ax$ are such kernels, and a whole further set is furnished by a theorem established by Hankel, namely[2]

$$g(x) = \int_0^\infty ds\,(xs)^{\frac{1}{2}}J_\nu(xs) \int_0^\infty g(x)(xs)^{\frac{1}{2}}J_\nu(xs)\,dx,$$

[2] Where $f(x)$ is discontinuous, the left-hand side should be replaced by $\frac{1}{2}[f(x+0) + f(x-0)]$.

whence $$G(s) = \int_0^\infty g(x)(xs)^{\frac{1}{2}} J_\nu(xs)\,dx,$$

and conversely, $$g(x) = \int_0^\infty G(s)(xs)^{\frac{1}{2}} J_\nu(xs)\,ds.$$

By splitting off a factor $s^{\frac{1}{2}}$ from $G(s)$ and $x^{\frac{1}{2}}$ from $g(x)$, that is, by putting $G(s) = s^{\frac{1}{2}}F(s)$ and $g(x) = x^{\frac{1}{2}}f(x)$, we obtain the following alternative expressions of the above formulas:

$$F(s) = \int_0^\infty xf(x)J_v(xs)\,dx$$

$$f(x) = \int_0^\infty sF(s)J_v(xs)\,ds.$$

The case in which $\nu = 0$ was derived earlier from the two-dimensional Fourier transform under conditions of circular symmetry.

It is interesting that by taking $\nu = \pm\frac{1}{2}$ and using the relations

$$J_{\frac{1}{2}}(z) = \left(\frac{2}{\pi z}\right)^{\frac{1}{2}} \sin z, \qquad J_{-\frac{1}{2}}(z) = \left(\frac{2}{\pi z}\right)^{\frac{1}{2}} \cos z,$$

we recover the known kernels $2 \cos 2\pi ax$ and $2 \sin 2\pi ax$, which shows that the cosine and sine transform formulas are included in Hankel's theorem.

THE THREE-DIMENSIONAL FOURIER TRANSFORM

Undoubtedly physical systems have three dimensions, but for reasons of theoretical tractability, one seeks simplifications. The classical example in which Fourier analysis in three dimensions has nevertheless had to be faced is the diffraction of X-rays by crystals. The formulas are

$$F(u,v,w) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y,z) e^{-i2\pi(xu+yv+zw)}\,dx\,dy\,dz$$

$$f(x,y,z) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(u,v,w) e^{i2\pi(ux+vy+wz)}\,du\,dv\,dw.$$

Multidimensional transforms, should they be encountered, will be recognized without difficulty. By taking $\mathbf{x}$ and $\mathbf{s}$ to be vectors whose components are $(x_1,x_2,\ldots)$ and $(s_1,s_2,\ldots)$, we have the following convenient vector notation for n-dimensional transforms:

$$F(\mathbf{s}) = \int\int \cdots \int_{-\infty}^{\infty} f(\mathbf{x}) e^{-i2\pi \mathbf{x}\mathbf{s}}\,dx_1\,dx_2 \ldots dx_n.$$

In cylindrical coordinates r, θ, z where

$$x + iy = re^{i\theta},$$

the three-dimensional transform may be expressed in terms of the transform variables s, ϕ, w, where

$$u + iv = se^{i\phi},$$

by the formulas

$$G(s,\phi,w) = \int_0^\infty \int_0^{2\pi} \int_{-\infty}^\infty g(r,\theta,z)e^{-i2\pi[sr\cos(\theta-\phi)+wz]}r\,dr\,d\theta\,dz$$

$$g(r,\theta,z) = \int_0^\infty \int_0^{2\pi} \int_{-\infty}^\infty G(s,\phi,w)e^{i2\pi[sr\cos(\theta-\phi)+wz]}s\,ds\,d\phi\,dw.$$

These results are derivable directly from the basic formulas by substituting

$$g(r,\theta,z) = f(x,y,z)$$

and $$G(s,\phi,w) = F(u,v,w).$$

Under circular symmetry, that is, when f is independent of θ (and hence F independent of ϕ), we find by writing

$$h(r,z) = f(x,y,z)$$

and $$H(s,w) = F(u,v,w)$$

that $$H(s,w) = 2\pi\int_0^\infty \int_{-\infty}^\infty h(r,z)J_0(2\pi sr)e^{-i2\pi wz}r\,dr\,dz$$

$$h(r,z) = 2\pi\int_0^\infty \int_{-\infty}^\infty H(s,w)J_0(2\pi sr)e^{i2\pi wz}s\,ds\,dw.$$

To obtain this result we use the formula derived earlier for the Hankel transform of zero order.

Under cylindrical symmetry, that is, when f is independent of both θ and z, being a function of r only, say $f(x,y,z) = \mathbf{k}(r)$ and $F(u,v,w) = K(s,w)$, then

$$K(s,w) = \mathbf{K}(s)\,\delta(w),$$

where $$\mathbf{K}(s) = 2\pi\int_0^\infty \mathbf{k}(r)J_0(2\pi sr)r\,dr.$$

In spherical coordinates r,θ,ϕ, with transform variables s,Θ,Φ, we have

$$x = r\sin\theta\cos\phi \qquad y = r\sin\theta\sin\phi \qquad z = r\cos\theta$$
$$u = s\sin\Theta\cos\Phi \qquad v = s\sin\Theta\sin\Phi \qquad w = s\cos\Theta.$$

Writing $f(x,y,z) = g(r,\theta,\phi)$ and $F(u,v,w) = G(s,\Theta,\Phi)$, we find

$$G(s,\Theta,\Phi) = \int_0^\infty \int_0^\pi \int_0^{2\pi} g(r,\theta,\phi)e^{-i2\pi sr[\cos\Theta\cos\theta+\sin\Theta\sin\theta\cos(\phi-\Phi)]}r^2\sin\theta\,dr\,d\theta\,d\phi.$$

$$g(r,\theta,\phi) = \int_0^\infty \int_0^\pi \int_0^{2\pi} G(s,\Theta,\Phi)e^{i2\pi sr[\cos\Theta\cos\theta+\sin\Theta\sin\theta\cos(\phi-\Phi)]}s^2\sin\Theta\,ds\,d\Theta\,d\Phi.$$

With circular symmetry, that is, when $f(x,y,z)$ is independent of ϕ, we have, writing $f(x,y,z) = h(r,\theta)$ and $F(u,v,w) = H(s,\Theta)$,

$$H(s,\Theta) = 2\pi\int_0^\infty\int_0^\pi h(r,\theta)J_0(2\pi sr \sin\Theta \sin\theta)e^{-i2\pi sr\cos\Theta\cos\theta}r^2 \sin\theta\, dr\, d\theta.$$

With spherical symmetry, we have, writing $f(x,y,z) = k(r)$ and

$$F(u,v,w) = K(s),$$

$$K(s) = 4\pi\int_0^\infty k(r)\,\mathrm{sinc}\,(2sr)\,r^2\,dr,$$

$$k(r) = 4\pi\int_0^\infty K(s)\,\mathrm{sinc}\,(2sr)\,s^2\,ds.$$

A few examples of three-dimensional Fourier transforms are given in Table 13.4. Many more can be generated by noting that

$$f(x)g(y)h(z) \supset F(u)G(v)H(w),$$

where $f(x)$, $F(u)$, and the like are one-dimensional Fourier transform pairs. This result is proved by expressing $f(x)g(y)h(z)$ in the form

$$f(x)\,\delta(y)\,\delta(z) * \delta(x)g(y)\,\delta(z) * \delta(x)\,\delta(y)h(z)$$

TABLE 13.4
Some three-dimensional Fourier transforms†

$f(x,y,z)$		$F(u,v,w)$
${}^3\delta(x-a, y-b, z-c)$	point	$e^{i2\pi(au+bv+cw)}$
$e^{-\pi(x^2/a^2+y^2/b^2+z^2/c^2)}$	Gaussian	$abce^{-\pi(a^2u^2+b^2v^2+c^2w^2)}$
${}^3\Pi(x,y,z)$	cube	$\mathrm{sinc}\,u\,\mathrm{sinc}\,v\,\mathrm{sinc}\,w$
${}^2\Pi(x,y)$	bar	$\mathrm{sinc}\,u\,\mathrm{sinc}\,v\,\delta(w)$
$\Pi(x)$	slab	$\mathrm{sinc}\,u\,\delta(v)\,\delta(w)$
$\Pi(x)\Pi[(y^2+z^2)^{\frac{1}{2}}]$	disk	$\mathrm{sinc}\,u\dfrac{J_1[\pi(v^2+w^2)^{\frac{1}{2}}]}{2(v^2+w^2)^{\frac{1}{2}}}$
$\Pi\left(\dfrac{r}{2}\right)$	ball	$\dfrac{\sin 2\pi s - 2\pi s\cos 2\pi s}{2\pi^2 s^3}$
$(1-\lvert r\rvert)\Pi\left(\dfrac{r}{2}\right)$		$\dfrac{\pi}{3}\dfrac{12}{(2\pi)^4}\dfrac{2(1-\cos 2\pi s) - 2\pi s\sin 2\pi s}{s^4}$
$(1-r^2)\Pi\left(\dfrac{r}{2}\right)$		$\dfrac{8\pi}{(2\pi)^5}\dfrac{[3-(2\pi s)^2]\sin 2\pi s - 3(2\pi s)\cos 2\pi s}{s^5}$
$\dfrac{e^{-r/R}}{\frac{4}{3}\pi R^3}$		$\dfrac{6}{(1+4\pi^2R^2s^2)^2}$
$e^{-\pi r^2}$		$e^{-\pi s^2}$

†In this table $r^2 = x^2 + y^2 + z^2$ and $s^2 = u^2 + v^2 + w^2$.

and then applying the convolution theorem in three dimensions. As various cases of this kind we have

$$\begin{aligned} f(x)g(y)h(z) &\supset F(u)G(v)H(w) \\ f(x)g(y) &\supset F(u)G(v)\,\delta(w) \\ f(x) &\supset F(u)\,\delta(v)\,\delta(w) \\ \mathbf{k}(r)h(z) &\supset \mathbf{K}(s)H(w) \\ \mathbf{k}(r) &\supset \mathbf{K}(s)\,\delta(w) \end{aligned}$$

THE HANKEL TRANSFORM IN n DIMENSIONS

When symmetry in n dimensions exists, the resulting one-dimensional transform is

$$\mathcal{F}(q) = \frac{2\pi}{q^{\frac{1}{2}n-1}} \int_0^\infty f(r) J_{\frac{1}{2}n-1}(2\pi q r) r^{\frac{1}{2}n}\, dr.$$

By putting $n = 1, 2, 3$, we recover the Fourier and Hankel transforms and the case of spherical symmetry previously stated (bearing in mind that $J_{\frac{1}{2}}(x) = (2/\pi x)^{\frac{1}{2}} \sin x$ and $J_{-\frac{1}{2}}(x) = (2/\pi x)^{\frac{1}{2}} \cos x$).

THE MELLIN TRANSFORM

The transform defined for complex s by

$$F_M(s) = \int_0^\infty f(x) x^{s-1}\, dx$$

proves to be equivalent to the Laplace transform (Chapter 14) when a simple change of variable is made. Putting

$$x = e^{-t}$$

we have

$$\begin{aligned} dx &= -e^{-t}\, dt \\ x^{s-1} &= e^{-t(s-1)}, \end{aligned}$$

whence

$$F_M(s) = \int_{-\infty}^{\infty} f(e^{-t}) e^{-st}\, dt.$$

The Laplace transform of the t functions, and the Mellin transform of the x function, shown in Fig. 13.5, are the same. When we replot a function of time as a function of e^{-t}, we compress all positive time into the range from 1 to 0.

Several Mellin transforms are listed in Table 13.5.

The inversion formula for the Mellin transform is

$$f(x) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} F_M(s) x^{-s}\, ds.$$

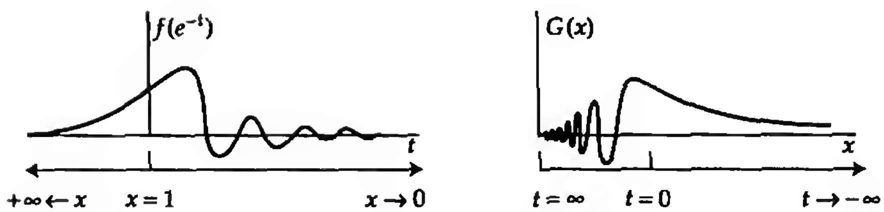


Fig. 13.5 Relation between Mellin and Laplace transforms.

Just as the Laplace transform approaches the Fourier series as a special case when the nonzero parts of $F_L(s)$ concentrate around imaginary integral values of s, so is the Mellin transform related to power series.

We may regard $F_M(s)$ as the $(s-1)$th moment of $f(x)$. For example, if $f(x) = \Pi(x - \frac{1}{2})$, then $F_M(s) = s^{-1}$ (provided Re s is positive). This Mellin transform pair is illustrated in Fig. 13.6. In general,

$$\begin{aligned}
&\text{Area of } f(x) &&= F_M(1)\\
&\text{First moment of } f(x) &&= F_M(2)\\
&\text{Second moment of } f(x) &&= F_M(3)\\
&\text{Abscissa of centroid, } \langle x \rangle &&= \frac{F_M(2)}{F_M(1)}\\
&\text{Radius of gyration} &&= \left[\frac{F_M(3)}{F_M(1)}\right]^{\frac{1}{2}}\\
&\text{Second moment about centroid} &&= F_M(3) - \frac{[F_M(2)]^2}{F_M(1)}
\end{aligned}$$

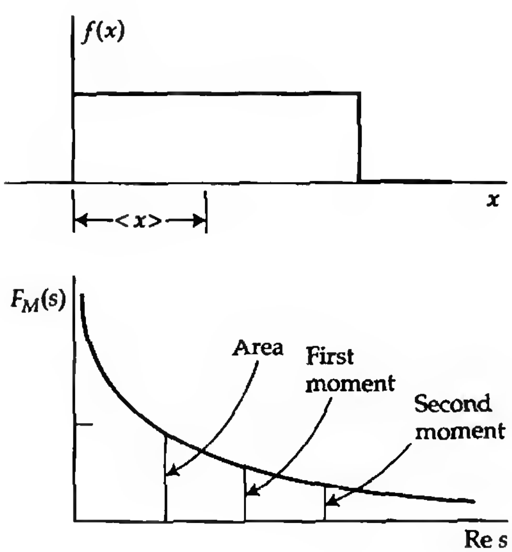


Fig. 13.6 Mellin transform pair.

TABLE 13.5
Some Mellin transforms

$f(x)$	$F_M(s)$	Convergence condition
$\delta(x-a)$	a^{s-1}	
$H(x-a)$	$-\dfrac{a^s}{s}$	
$H(a-x)$	$\dfrac{a^s}{s}$	
$x^n H(x-a)$	$-\dfrac{a^{s+n}}{s+n}$	
$x^n H(a-x)$	$\dfrac{a^{s+n}}{s+n}$	
e^{-ax}	$a^{-s}\Gamma(s)$	Re $a > 0$, Re $s > 0$
e^{-x^2}	$\frac{1}{2}\Gamma(\frac{1}{2}s)$	
$\sin x$	$\Gamma(s)\sin\frac{1}{2}\pi s$	$-1 <$ Re $s < 1$
$\cos x$	$\Gamma(s)\cos\frac{1}{2}\pi s$	$0 <$ Re $s < 1$
$\dfrac{1}{1+x}$	$\pi \operatorname{cosec} \pi s$	
$\dfrac{1}{1-x}$	$\pi \cot \pi s$	
$\dfrac{1}{(1+x)^a}$	$\dfrac{\Gamma(s)\Gamma(a-s)}{\Gamma(a)}$	Re $a > 0$
$\dfrac{1}{1+x^2}$	$\frac{1}{2}\pi \operatorname{cosec} \frac{1}{2}\pi s$	
$(1-x)^{a-1}H(1-x)$	$\dfrac{\Gamma(s)\Gamma(a)}{\Gamma(s+a)}$	Re $a > 0$
$(x-1)^a H(x-1)$	$\dfrac{\Gamma(a-s)\Gamma(1-a)}{\Gamma(1-s)}$	$0 <$ Re $a < 1$
$\ln(1+x)$	$\dfrac{\pi}{s}\operatorname{cosec}\pi s$	$-1 <$ Re $s < 0$
$\frac{1}{2}\pi - \tan^{-1} x$	$\frac{1}{2}\pi s^{-1} \sec \frac{1}{2} s$	
$\Lambda(x-1)$	$\begin{cases} \dfrac{2(2^s-1)}{s(s+1)} & s \neq 0 \\ 2\ln 2 & s = 0 \end{cases}$	Re $s > -1$
$\operatorname{erfc} x$	$\dfrac{\Gamma(\frac{1}{2}s+\frac{1}{2})}{\pi^2 s}$	Re $s > 0$
$\operatorname{Si} x$	$-\dfrac{\Gamma(s)\sin\frac{1}{2}\pi s}{s}$	$-1 <$ Re $s < 0$

TABLE 13.6
Theorems for the Mellin transform

$f(x)$	$F_M(s)$	
$f(ax)$	$a^{-s}F_M(s)$	$a > 0$
$f_1 + f_2$	$F_{1M} + F_{2M}$	
$xf(x)$	$F_M(s+1)$	
$x^a f(x)$	$F_M(s+a)$	
$f\left(\frac{1}{x}\right)$	$F_M(-s)$	
$f(x^a)$	$a^{-1}F_M\left(\frac{s}{a}\right)$	$a > 0$
$f(x^{-a})$	$a^{-1}F_M\left(-\frac{s}{a}\right)$	$a > 0$
$f'(x)$	$-(s-1)F_M(s-1)$	
$xf'(x)$	$-sF_M(s)$	
$x^a f'(x)$	$-(s+a-1)F_M(s+a^{-1})$	
$f''(x)$	$(s-1)(s-2)F_M(s-2)$	
$x^2 f''(x)$	$s(s+1)F_M(s)$	
$f^{(n)}(x)$	$(-1)^n(s-n)\ldots(s-1)F_M(s-n)$	
$\left(x\frac{d}{dx}\right)^n f(x)$	$(-1)^n s^n F_M(s)$	
$\int_0^\infty f\left(\frac{x}{u}\right)g(u)\frac{du}{u}$	$F_M(s)G_M(s+1)$	
$\int_0^\infty f\left(\frac{x}{u}\right)g(u)\,du$	$F_M(s)G_M(s)$	
$f(x)g(x)$	$\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} F(\tau)G(s-\tau)\,d\tau$	

Theorems for the Mellin transform are listed in Table 13.6. These theorems are closely related to those discussed earlier in respect to Fourier transforms.

By taking Mellin transforms of each of the equations

$$g(\alpha) = \int_0^\infty f(x)k(\alpha x)\,dx$$

and

$$f(x) = \int_0^\infty g(\alpha)h(\alpha x)\,d\alpha,$$

the relation between the Mellin transforms $K_M(s)$ and $H_M(s)$ is found to be

$$K_M(s)H_M(1-s) = 1.$$

This relation leads to the solving kernel $h(\alpha x)$ for integral equations with kernel $k(\alpha x)$ and also, when $h(\alpha x) = k(\alpha x)$, leads to the condition for Fourier kernels of the form $k(\alpha x)$, namely,

$$K_M(s)K_M(1-s) = 1.$$

THE z TRANSFORM

A good deal of interest attaches to signals whose sample values at equispaced intervals of time constitute the full available information about the signal. Such signals arise in communications systems using pulse modulation, where several different messages are interlaced on a time-sharing basis, and another important sphere of interest includes control systems in which feedback is applied on a basis of samples, taken at regular intervals, of some quantity which is to be controlled. Many other examples come to mind, including topics involving probability theory and periodic structures in space, but pulse modulation and sampled-data control systems are the main fields where the z transform is customarily introduced.

It is usual to suppose that the time interval between samples is T, but here the interval will be taken to be unity. If it is necessary to deal with some other interval, one can always introduce a new independent variable differing by a suitable factor.

Let the function $f(t)$ be known to us only through its sample values at $t = 0, 1, 2, \ldots, n$, namely,

$$f(0) \quad f(1) \quad f(2) \ldots f(n).$$

Then the polynomial

$$F(z) = f(0) + f(1)z^{-1} + f(2)z^{-2} + \ldots + f(n)z^{-n}$$

is referred to as the z transform of $f(t)$. The function $f(t)$ is real and is taken to be zero for $t < 0$; the variable z is taken to be complex.

The impulse string

$$\phi(t) = f(0)\,\delta(t) + f(1)\,\delta(t-1) + f(2)\,\delta(t-2) + \ldots + f(n)\,\delta(t-n),$$

though quite different in character from the original function $f(t)$, is fully equivalent to the set of samples, as has been pointed out in connection with the sampling theorem. In fact it is often convenient to deal directly with the sequence

$$\{f(0) \quad f(1) \quad f(2) \cdots f(n)\}$$

without specifying the kind of signal.

If such signals are applied to the input of a system whose impulse response is

$$\{h(0) \quad h(1) \quad h(2) \ldots h(n)\},$$

then the output of the system is represented by the serial product

$$\{f(0)\ f(1)\ f(2)\ \ldots\ f(n)\} * \{h(0)\ h(1)\ h(2)\ \ldots\ h(n)\}.$$

But we have seen in Chapter 3 that the serial-product rule for convolving sequences is the same as that for multiplying polynomials. Hence the z transform of the output is the product

$$F(z)H(z),$$

where $H(z)$ is the z transform of the impulse response.

For example, if a signal $\{2\ \ 1\}$ is applied to a system whose impulse response is $\{8\ \ 4\ \ 2\ \ 1\}$, then

$$\begin{aligned} F(z) &= 2 + z^{-1} \\ H(z) &= 8 + 4z^{-1} + 2z^{-2} + z^{-3} \end{aligned}$$

and the z transform of the output is given by

$$\begin{aligned} F(z)H(z) &= (2 + z^{-1})(8 + 4z^{-1} + 2z^{-2} + z^{-3}) \\ &= 16 + 16z^{-1} + 8z^{-2} + 4z^{-3} + z^{-4}. \end{aligned}$$

Hence the output sequence is

$$\{16\ \ 16\ \ 8\ \ 4\ \ 1\}.$$

Conversely, if the input and output are given and it is desired to find what impulse response the system must have, one simply divides the z transform of the output by the z transform of the input and expands the result as a polynomial in z^{-1}. The coefficients then give the answer.

Serial multiplication and its inversion, as introduced in Chapter 3, constitute an alternative and quite direct approach to the theory of such numerical problems, and present the essential arithmetic of multiplying and dividing z transforms which remains when the numerous z's are omitted.

The possibility of factoring polynomials entering into products or quotients makes for theoretical convenience of the z transform and allows one's familiarity with algebraic manipulation to be brought into play. For this reason, a table of z transforms of some common impulse signals is useful for reference. If the signal is specified by a sequence of numerical data, it is a trivial matter to write down the z transform. We consider here cases in which the strengths of the successive impulses are given by some simple formula instead of by actual data. Let

$$\phi(t) = f(0)\,\delta(t) + f(1)\,\delta(t-1) + f(2)\,\delta(t-2) + \ldots + f(n)\,\delta(t-n).$$

Taking the Laplace transform of $\phi(t)$ and calling it $F_L(p)$, we have

$$F_L(p) = \int_{-\infty}^{\infty} \phi(t)e^{-pt}\,dt = \sum_{n=0}^{\infty} f(n)e^{-np}.$$

This is a simple polynomial in $\exp(-p)$. Putting

$$z = e^p,$$

we find
$$F_L(p) = \sum_{n=0}^{\infty} f(n)z^{-n},$$

but this polynomial is by definition the z transform of $f(t)$.

As an example, consider the sequence of pulses $\delta(t) + \delta(t-1) + \delta(t-2) \ldots$. Its Laplace theorem is given by $1 + e^{-p} + e^{-2p} + \ldots$, and its z transform is $1 + z^{-1} + z^{-2} + \ldots$. Thus the z transform is directly deducible from the Laplace transform. In this case of a function specifiable by a simple formula rather than by data, the polynomial can be expressed in finite form,

$$\frac{1}{1-e^{-p}} = \frac{1}{1-z^{-1}},$$

and makes a compact entry in Table 13.7 where z transforms are listed.

Operations carried out on $f(n)$ will correspond with operations on the z transform; pairs of such operations are listed in Table 13.8.

A comparison of this table with the theorems quoted for the Mellin transform will reveal that the two are virtually identical. This is because the Mellin transform and the inverse z transform are related to the Laplace transform in much the same way. Thus in the Laplace integral

$$\int_{-\infty}^{\infty} f(t)e^{-pt}\,dt$$

we put $x = \exp(-t)$ to derive the Mellin transform, while we put $z = \exp(-p)$ to derive the z transform. The three transforms and their inversion integrals are listed below for reference and comparison.

$$F_L(p) = \int_{-\infty}^{\infty} f(t)e^{-pt}\,dt \qquad f(t) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} F_L(p)e^{pt}\,dp$$

$$F_M(s) = \int_0^{\infty} f(x)x^{s-1}\,dx \qquad f(x) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} F_M(s)x^{-s}\,ds$$

$$f(n) = \frac{1}{2\pi i}\int_{\Gamma} F(z)z^{n-1}\,dz \qquad F(z) = \int_{-\infty}^{\infty} \phi(t)z^{-t}\,dt$$

$$= \sum_0^{\infty} f(n)z^{-n}$$

It is clear that the z transform is like the inverse Mellin transform except that t must assume real values whereas s may be complex, and conversely, x is real whereas z may be complex. The contour Γ on the z plane may be understood as follows. It must enclose the poles of the integrand. If the contour $c - i\infty$ to $c + i\infty$ for inverting the Laplace transformation is chosen to the right of all poles, then the circle into which it is transformed by the transformation $z = \exp(-p)$ will enclose all poles. In the common case where $c = 0$ is suitable (all poles of $F_L(p)$ in the left half-plane), the contour Γ becomes the circle $|x| = 1$.

TABLE 13.7
Some z transforms for algebraically specifiable signals

$\{f(n)\}$	$f(n)$	$F(z)$
$\{a_0\,a_1\,a_2\,\ldots\}$	a_n	$a_0 + a_1 z^{-1} + a_2 z^{-2} + \ldots$
$\{1\,1\,1\,1\,\ldots\}$	1	$\dfrac{1}{1 - z^{-1}}$
$\{0\,1\,1\,1\,\ldots\}$		$\dfrac{z^{-1}}{1 - z^{-1}}$
$\{0\,1\,2\,3\,\ldots\}$	n	$\dfrac{z^{-1}}{(1 - z^{-1})^2}$
$\{0\,1\,4\,9\,\ldots\}$	n^2	$\dfrac{z^{-1}(1 + z^{-1})}{(1 - z^{-1})^3}$
$\{0\,1\,8\,27\,\ldots\}$	n^3	$\dfrac{3z^{-2}(1 + z^{-1})}{(1 - z^{-1})^4} + \dfrac{z^{-1}(1 + 2z^{-1})}{(1 - z^{-1})^3}$
$\{1\,e^{-\alpha}\,e^{-2\alpha}\,\ldots\}$	$e^{-\alpha n}$	$\dfrac{1}{1 - e^{-\alpha}z^{-1}}$
$\{0\;\; 1 - e^{-\alpha}\;\; 1 - e^{-2\alpha}\,\ldots\}$	$1 - e^{-\alpha n}$	$\dfrac{(1 - e^{-\alpha})z^{-1}}{(1 - z^{-1})(1 - e^{-\alpha}z^{-1})}$
$\{0\,e^{-\alpha}\,2e^{-2\alpha}\,\ldots\}$	$ne^{-\alpha n}$	$\dfrac{e^{-\alpha}z^{-1}}{(1 - e^{-\alpha}z^{-1})^2}$
$\{0\,e^{-\alpha}\,4e^{-2\alpha}\,\ldots\}$	$n^2e^{-\alpha n}$	$\dfrac{e^{-\alpha}z^{-1}(1 + e^{-\alpha}z^{-1})}{(1 - e^{-\alpha}z^{-1})^3}$
	$\dfrac{1}{\alpha}(\alpha n - 1 + e^{-\alpha n})$	$\dfrac{z^{-1}}{(1 - z^{-1})^2} - \dfrac{(1 - e^{-\alpha})z^{-1}}{\alpha(1 - z^{-1})(1 - e^{-\alpha}z^{-1})}$
	$e^{-\alpha n} - e^{-\beta n}$	$\dfrac{(e^{-\alpha} - e^{-\beta})z^{-1}}{(1 - e^{-\alpha}z^{-1})(1 - e^{-\beta}z^{-1})}$
	$\sin \alpha n$	$\dfrac{z^{-1}\sin\alpha}{1 - z^{-1}2\cos\alpha + z^{-2}}$
	$\cos \alpha n$	$\dfrac{1 - z^{-1}\cos\alpha}{1 - z^{-1}2\cos\alpha + z^{-2}}$
	$\sinh \alpha n$	$\dfrac{z^{-1}\sinh\alpha}{1 - z^{-1}2\cosh\alpha + z^{-2}}$
	$\cosh \alpha n$	$\dfrac{1 - z^{-1}\cosh\alpha}{1 - z^{-1}2\cosh\alpha + z^{-2}}$
	$e^{-\alpha n}\cos\omega n$	$\dfrac{1 - z^{-1}e^{-\alpha}\cos\omega}{1 - z^{-1}2e^{-\alpha}\cos\omega + e^{-2\alpha}z^{-2}}$
	$e^{-\alpha n}\sin\omega n$	$\dfrac{z^{-1}e^{-\alpha}\sin\omega}{1 - z^{-1}2e^{-\alpha}\cos\omega + e^{-2\alpha}z^{-2}}$
	$f(n+1)$	$zF(z) - f(0)z$
	$f(n+2)$	$z^2F(z) - f(0)z^2 - f(1)z$
	$n^a f(n)$	$-z(d/dz)^a F(z)$

TABLE 13.8
Theorems for the z transform

$f(n)$	$F(z)$	
$f_1(n) + f_2(n)$	$F_1(z) + F_2(z)$	
$f(n-1)$	$z^{-1}F(z)$	
$f(n-a)$	$z^{-a}F(z)$	$a \geq 0$†
$nf(n)$	$-zF'(z)$	
$-(n-1)f(n-1)$	$F'(z)$	
$a^{-n}f(n)$	$F(az)$	$a > 0$
$f_1(n) * f_2(n)$	$F_1(z)F_2(z)$	

†This restriction is a consequence of assuming $f(n)$ to be zero for $n < 0$.

THE ABEL TRANSFORM

As soon as one goes beyond the one-dimensional applications of Fourier transforms and into optical-image formation, television-raster display, mapping by radar or passive detection, and so on, one encounters phenomena which invite the use of the Abel transform for their neatest treatment. These phenomena arise when circularly symmetrical distributions in two dimensions are projected in one dimension. A typical example is the electrical response of a television camera as it scans across a narrow line; another is the electrical response of a microdensitometer whose slit scans over a circularly symmetrical density distribution on a photographic plate.

Fractional-order derivatives are also closely connected with the Abel transform, which therefore also arises in fields, such as conduction of heat in solids or transmission of electrical signals through cables, where fractional-order derivatives are encountered.

The Abel transform $f_A(x)$ of the function $f(r)$ is commonly defined as

$$f_A(x) = 2\int_x^\infty \frac{f(r)r\,dr}{(r^2 - x^2)^{\frac{1}{2}}}.$$

The choice of the symbols x and r is suggested by the many applications in which they represent an abscissa and a radius, respectively, in the same plane.

The above formula may be written

$$f_A(x) = \int_0^\infty k(r,x)f(r)\,dr,$$

where
$$k(r,x) = \begin{cases} 2r(r^2 - x^2)^{-\frac{1}{2}} & r > x \\ 0 & r < x. \end{cases}$$

The kernel $k(r,x)$, regarded as a function of r in which x is a parameter, shifts to the right as x increases, and it also changes its form. A slight change of variable leads to a kernel which simply shifts without change of form. Thus putting $\xi = x^2$ and $\rho = r^2$, and letting $f_A(x) = F_A(x^2)$ and $f(r) = F(r^2)$, we have

$$F_A(\xi) = \int_0^\infty K(\xi - \rho)F(\rho)\,d\rho,$$

where
$$K(\xi) = \begin{cases} (-\xi)^{-\frac{1}{2}} & \xi < 0 \\ 0 & \xi \geq 0; \end{cases}$$

alternatively,
$$F_A(\xi) = \int_\rho^\infty \frac{F(\rho)\,d\rho}{(\rho - \xi)^{\frac{1}{2}}}$$

or again,
$$F_A = K * F.$$

When necessary, F_A will be referred to as the "modified Abel transform of F." Having reduced the formula to a convolution integral, we may take Fourier transforms and write

$$\overline{F}_A = \overline{K}\,\overline{F}.$$

Since
$$\overline{K}(s) = \frac{1}{(-2is)^{\frac{1}{2}}}$$

it follows that
$$\overline{F} = (-2is)^{\frac{1}{2}}\overline{F}_A$$

$$= -\frac{1}{\pi}\frac{1}{(-2is)^{\frac{1}{2}}}\,i2\pi s\overline{F}_A$$

whence
$$F = -\frac{1}{\pi}K * F_A';$$

that is,
$$F(\rho) = -\frac{1}{\pi}\int_\rho^\infty \frac{F_A'(\xi)\,d\xi}{(\xi - \rho)^{\frac{1}{2}}}.$$

The solution of the modified Abel integral equation enables F to be expressed in terms of the derivative of F_A. Integrating the solution by parts, or choosing different factors for the transform of F, we obtain a solution in terms of the second derivative of F_A:

$$F = \frac{2}{\pi}\mathcal{H} * F_A''$$

where

$$\mathcal{H}(\xi) = \begin{cases} (-\xi)^{\frac{1}{2}} & \xi < 0 \\ 0 & \xi \geq 0. \end{cases}$$

Since $\overline{K}$ is nowhere zero, the solution is unique (except for additive null functions).

Reverting to f and f_A, we may write the solutions as

$$f(r) = -\frac{1}{\pi}\int_r^\infty \frac{f'_A(x)\,dx}{(x^2 - r^2)^{\frac{1}{2}}} = -\frac{1}{\pi}\int_r^\infty (x^2 - r^2)^{\frac{1}{2}} \frac{d}{dx}\left[\frac{f'_A(x)}{x}\right] dx,$$

or, if the integral is zero beyond $x = r_0$, and allowing for the possibility that the integrand may behave impulsively at r_0, we have

$$\begin{aligned} f(r) &= -\frac{1}{\pi}\int_r^{r_0} \frac{f'_A(x)\,dx}{(x^2 - r^2)^{\frac{1}{2}}} + \frac{f_A(r_0)}{\pi(r_0^2 - r^2)^{\frac{1}{2}}} \\ &= -\frac{1}{\pi}\int_r^{r_0} (x^2 - r^2)\frac{d}{dx}\left[\frac{f'_A(x)}{x}\right] dx - \frac{f'_A(r_0)}{\pi r_0}(r_0^2 - r^2)^{\frac{1}{2}}. \end{aligned}$$

Useful relations for checking Abel transforms are

$$\int_{-\infty}^{\infty} f_A(x)\,dx = 2\pi\int_0^\infty f(r)r\,dr$$

and

$$f_A(0) = 2\int_0^\infty f(r)\,dr.$$

Another property is that

$$K * K * F' = -\pi F;$$

that is, the operation $K\,*$ applied twice in succession annuls differentiation; then F_A is the half-order integral of F, and conversely, F is the half-order differential coefficient of F_A. To prove this, note that if $F_A = K * F$ implies that $F = -\pi^{-1}K * F'_A$, then it follows further that $F'_A = K * F'$; whence

$$K * K * F' = K * F'_A = -\pi F.$$

In Table 13.9 the first eight examples are to be taken as zero for r and x greater than a.

Numerical evaluation of Abel transforms is comparatively simple in view of the possibility of conversion to a convolution integral. One first makes the change of variable, then evaluates sums of products of $K(\rho)$ and $f(\xi - \rho)$ at discrete intervals of ρ. The values of K turn out to be the same, however fine an interval is chosen, save for a normalizing factor; consequently, a universal table of values (see Table 13.10) can be set up for permanent reference. The table shows coefficients for immediate use with values of F read off at $\rho = \frac{1}{2}, 1\frac{1}{2}, \ldots, 9\frac{1}{2}$, the scale of ρ being such that F becomes zero or negligible at $\rho = 10$. The table gives mean values of K over the intervals $0 - 1, 1 - 2, \ldots$. Thus at $\rho = n + \frac{1}{2}$ the value is

$$\int_n^{n+1} K(-\rho)\,d\rho = 2(n + 1)^{\frac{1}{2}} - 2n^{\frac{1}{2}}.$$

TABLE 13.9
Some Abel transforms

$f(r)$		$f_A(x)$	
$\Pi(r/2a)$	Disk	$2(a^2 - x^2)^{\frac{1}{2}}\Pi(x/2a)$	Semiellipse
$(a^2 - r^2)^{-\frac{1}{2}}\Pi(r/2a)$		$\pi\Pi(x/2a)$	Rectangle
$(a^2 - r^2)^{\frac{1}{2}}\Pi(r/2a)$	Hemisphere	$\frac{1}{2}\pi(a^2 - x^2)\Pi(x/2a)$	Parabola
$(a^2 - r^2)\Pi(r/2a)$	Paraboloid	$\frac{4}{3}(a^2 - x^2)^{\frac{3}{2}}\Pi(x/2a)$	
$(a^2 - r^2)^{\frac{3}{2}}\Pi(r/2a)$		$(3\pi/8)(a^2 - x^2)^2\Pi(x/2a)$	
$a\Lambda(r/a)$	Cone	$[a(a^2 - x^2)^{\frac{1}{2}} - x^2\cosh^{-1}(a/x)]\Pi(x/2a)$	
$\pi^{-1}\cosh^{-1}(a/r)\Pi(r/2a)$		$a\Lambda(x/a)$	Triangle
$\delta(r - a)$	Ring impulse	$2a(a^2 - x^2)^{-\frac{1}{2}}\Pi(x/2a)$	
$\exp(-r^2/2\sigma^2)$	Gaussian	$(2\pi)^{\frac{1}{2}}\sigma\exp(-x^2/2\sigma^2)$	Gaussian
$r^2\exp(-r^2/2\sigma^2)$		$(2\pi)^{\frac{1}{2}}\sigma(x^2 + \sigma^2)\exp(-x^2/2\sigma^2)$	
$(r^2 - \sigma^2)\exp(-r^2/2\sigma^2)$		$(2\pi)^{\frac{1}{2}}\sigma x^2\exp(-x^2/2\sigma^2)$	
$(a^2 + r^2)^{-1}$		$\pi(a^2 + x^2)^{-\frac{1}{2}}$	
$J_0(2\pi ar)$		$(\pi a)^{-1}\cos 2\pi ax$	
$2\pi\left[r^{-3}\int_0^r J_0(r)\,dr - r^{-2}J_0(r)\right] = M(r)$		$\mathrm{sinc}^2 x$	
$\delta(r)/\pi\lvert r\rvert$		$\delta(x)$	
$2a\,\mathrm{sinc}\,2ar$		$J_0(2\pi ax)$	
$\frac{1}{2}r^{-1}J_1(2\pi ar)$		$\mathrm{sinc}\,2ax$	

TABLE 13.10
Coefficients for performing or inverting the Abel transformation

ρ	K	ρ	K	ρ	K	ρ	K
$\frac{1}{2}$	2.000	$5\frac{1}{2}$	0.427	$10\frac{1}{2}$	0.309	$15\frac{1}{2}$	0.254
$1\frac{1}{2}$	0.828	$6\frac{1}{2}$	0.393	$11\frac{1}{2}$	0.295	$16\frac{1}{2}$	0.246
$2\frac{1}{2}$	0.636	$7\frac{1}{2}$	0.364	$12\frac{1}{2}$	0.283	$17\frac{1}{2}$	0.239
$3\frac{1}{2}$	0.536	$8\frac{1}{2}$	0.343	$13\frac{1}{2}$	0.272	$18\frac{1}{2}$	0.233
$4\frac{1}{2}$	0.472	$9\frac{1}{2}$	0.325	$14\frac{1}{2}$	0.263	$19\frac{1}{2}$	0.226

When N points of subdivision are used, the scale of ρ is arranged so that F becomes zero at $\rho = N$. The coefficients may then all be multiplied by $(10/N)^{\frac{1}{2}}$, or the coefficients may be left unchanged and the answers multiplied by $(10/N)^{\frac{1}{2}}$.

As an example consider $F(\rho) = (10 - \rho)^{\frac{1}{2}}$, for which the modified Abel transform is known to be $F_A(\xi) = \frac{1}{2}\pi(1 - \xi)$. We work at unit intervals and, visualizing the algorithm as a manual procedure, copy the coefficients on a movable strip. The calculation in progress is shown in Fig. 13.7. The movable strip is in position for calculating $F_A(\xi)$ as the sum of products of corresponding values of F and K:

$$7.78 = 2.12 \times 2.000 + 1.87 \times 0.828 + \dots + 0.71 \times 0.472.$$

The inverse problem, that of calculating F from F_A, can be handled by means of the relation $F = -\pi^{-1}K * F'_A$ if F_A is first differentiated. However, it will be perceived that the calculation just described can be done in reverse, using the values of F_A, and working the movable strip upward from the bottom. The strip is shown in position for calculating $F(5 - \frac{1}{2})$, let us say by means of a pocket calculator. Form the products $0.71 \times 0.472, \dots, 1.87 \times 0.828$, allowing them to accumulate in the memory. Subtract this sum of products from 7.78 and divide by 2.000 to obtain the next wanted value, $F(5 - \frac{1}{2}) = 2.12$.

These convolution procedures are readily programmable; an alternative approach, sharing the virtue of avoiding the singularity as r approaches x, is as follows. Set up a user-defined function **FNf(r)** using an algebraic expression or, if the radial function is available as data samples, make values of $f(r)$ between samples

ρ	F	K	F_A
			15.65
	3.08		
1			14.08
	2.91		
2			12.52
	2.74		
3			10.94
	2.55		
4			9.37
	2.35		
5			7.78
	2.12	2.000	
6			6.20
	1.87	0.828	
7			4.62
	1.58	0.636	
8			3.03
	1.22	0.536	
9			1.42
	0.71	0.472	
10			0
		0.427	

Fig. 13.7 Calculating modified Abel transforms.

available by interpolation. Here is the program, based on N equispaced integer values of $f(r)$ from 0 to $N-1$. For higher precision, increase M.

```
N=10
M=10
FOR i=0 TO N-1
   s=0.5*FNf(i/N)
   FOR j=1 TO INT(M*SQR(N^2-i^2))
      r=SQR(i^2+(j/M)^2)/N
      s=s+FNf(r)
   NEXT j
   PRINT 1/N;FNr(2*s/N/M)
NEXT i
```

An example starting from $f(r) = 1 - r$ gave these results. The mean absolute error is less than 10^{-5}.

r	0	.1	.2	.3	.4	.5	.6	.7	.8	.9
$f(r)$	1	.9	.8	.7	.6	.5	.4	.3	.2	.1
$f_A(x)$	1	.9651	.8881	.7853	.6658	.5368	.4045	.2753	.1564	.0575

THE RADON TRANSFORM AND TOMOGRAPHY

When a slit scans over a circularly symmetrical two-dimensional density distribution on a photographic plate, generating a profile $f_A(x)$ from a radial density function $f(r)$, we have one example of a way in which the one-dimensional Abel transform arises. If the density distribution is $f(x,y)$, which is not symmetrical but depends on two coordinates, the scans may still be taken but they will depend on the direction of scanning θ. Calling the abscissa for each scan R, we define the Radon transform $g_\theta(R)$ of the function $f(x,y)$ by

$$g_\theta(R) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\delta(R - x\cos\theta - \sin\theta)dxdy.$$

How to invert this transformation, that is, given the scans $g_\theta(R)$, usually for continuous R and a discrete set of scanning directions θ, to arrive at the unknown function $f(x,y)$, became very important with the advent of computer-assisted tomography. Chest x-rays of the lungs are degraded by unwanted images of ribs superposed both in front of and behind the features of interest. The possibility of obtaining the density distribution within a slice of tissue inside the body by x-rays without shadowing by ribs or other contiguous tissue has revolutionized diagnosis, first by computer-assisted tomography (CAT scans) and then by NMR (nuclear magnetic resonance) imaging. Tomography, before computing, was a technique of defocusing the foreground and background by translating the x-ray source during the exposure and simultaneously moving the x-ray plate in the opposite direction at the appropriate speed.

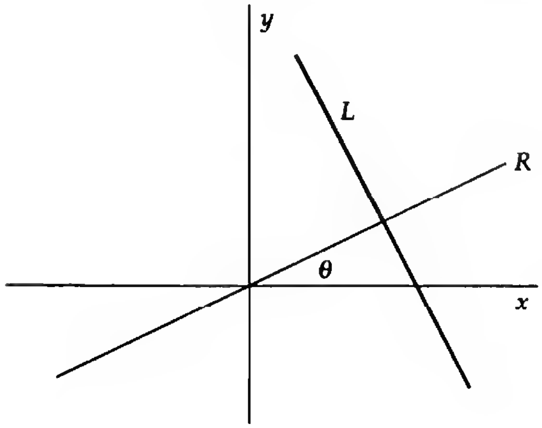


Fig. 13.8 **A density distribution in the (x,y)-plane is scanned by a line L moving in the R-direction to generate a projection.**

The factor $\delta(R - x \cos \theta - y \sin \theta)$ is zero everywhere except where its argument is zero, which is along the straight line $x \cos\theta + y \sin \theta = R$ (Fig. 13.8). This straight line L represents the slit when it is at a perpendicular distance R from the origin and inclined at an angle θ to the y-axis. If θ is kept fixed, say at a value θ_1, while R is varied, then the integral $g_{\theta_1}(R)$ constitutes the projection of the density distribution $f(x,y)$ onto the line $\theta = \theta_1$ as a function of R. The resulting profile is referred to as a single scan.

The practical computational method for inversion was arrived at by Fourier transforming the Radon integral equation, finding a method of solution, and then retransforming the steps to end up with data-plane operations in which numerical Fourier transformation is actually dispensed with. The technique, known as modified back-projection (Bracewell and Riddle, 1967), was developed in connection with radioastronomical imaging where a distributed source of radiation is scanned by an antenna that receives from a narrow strip of sky whose orientation θ can be varied between scans.

The inversion procedure derives from a remarkable connection that exists between the Fourier, Abel, and Hankel transforms and from a generalization known as the Projection-Slice Theorem.

The Abel-Fourier-Hankel ring of transforms. Starting with an even function $f(r)$, if we take the Abel transform, then take the Fourier transform, and finally take the Hankel transform, we return to the original function $f(r)$ as shown in Bracewell (1956). For example, starting with $f(r) = \delta(r - a)$, which is a ring impulse located on the circle $r = a$, we take the Abel transform (Table 13.9) to get $2a/\sqrt{a^2 - x^2}\Pi(x/2a)$, the Fourier transform of which is $2\pi a J_0(2\pi as)$ (Pictorial Dictionary). From Table 13.2 we verify that the Hankel transform of the Bessel function is $\delta(r - a)$, the function we started with.

Projection-slice theorem. When a two-dimensional density distribution is a function of radius alone, all three of the above transformations are one-dimensional but $f(r)$ can be generalized to become a function $f(x,y)$ of both x and y and

what was the Hankel transform above generalizes to a function $F(u,v)$ of u and v. These two two-dimensional functions constitute a two-dimensional Fourier transform pair, as explained earlier in the chapter in connection with the Hankel transformation. One way of thinking about Fourier transformation in two dimensions is to note that $F(u,0)$, the slice through $F(u,v)$ along the u-axis, is given by putting $v = 0$ in the two-dimensional Fourier transform definition to get

$$F(u,0) = \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} f(x,y)dy\right]e^{-i2\pi ux}dx.$$

The item in square brackets is the projection of $f(x,y)$ on the x-axis. The remaining integral with respect to x simply transforms the projection. In consequence, when $f(x, y)$ is given, one slice through $F(u,v)$, namely the one along the u-axis, is obtainable by first projecting $f(x,y)$ onto its x-axis and then taking a one-dimensional Fourier transform. The Projection-Slice Theorem says that the slice through $F(u,v)$ at any angle θ_1 in the (u,v)-plane, i.e., along a line parallel to the axis R in the (x,y)-plane, is obtainable as the Fourier transform of the projection of $f(x,y)$ onto the axis R in the (x,y)-plane (Bracewell, 1956).

Reconstruction by modified back projection. Now the process of tomography is to project a certain $f(x,y)$ at various angles θ, preferably numerous and equispaced; consequently those parts of the transform $F(u,v)$ can be deduced that lie on slices at corresponding angles. From knowledge of $F(u,v)$ one can recover $f(x,y)$ by two-dimensional Fourier transformation; but to do this one must first interpolate onto a square grid in the (u,v)-plane in order to be able to utilize available algorithms. Such numerical interpolation proves to take more time than the transformation. To avoid interpolation we note that in the (u,v)-plane, the data points resulting from the various one-dimensional transformations lie on diverging spokes $\theta =$ const. The density of points is thus inversely proportional to radius, a nonuniformity that can be corrected for by multiplication by the absolute value of radius in the (u,v)-plane. Let M be a spatial frequency in the (u,v)-plane beyond which no content is expected, and let $q = \sqrt{u^2 + v^2}$. Then the correction factor is $\Pi(q/2M) - \Lambda(q/M)$. After such correction a two-dimensional Fourier transform would deliver the desired $f(x,y)$.

But the multiplicative correction to values along the slice in the (u,v)-plane corresponds to a rather simple convolution operation on the original projections $g_\theta(R)$ in the data domain, an operation that produces a modified scan

$$\hat{g}_\theta(R) = g_\theta(R) * (2M \operatorname{sinc} 2M\,R - M \operatorname{sinc}^2 M\,R).$$

Thus the inversion procedure for the Radon transform is (a) to modify each measured scan by simple convolution to get $\hat{g}_\theta(R)$, (b) to back-project, and (c) to accumulate the separate back projections over the (x,y)-plane. Back projection is to distribute the modified scan $\hat{g}_{\theta_1}(R)$ uniformly over the (x,y)-plane in the direction perpendicular to the R-axis. For more details see Bracewell (1995) and Deans (1983).

THE HILBERT TRANSFORM

The Hilbert transform has a variety of applications; the ones to be mentioned here concern causality and the generalization of the phasor idea beyond pure alternating current.

We define the Hilbert transform of $f(x)$ by

$$F_{\mathrm{Hi}}(x) = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{f(x')\,dx'}{x' - x}.$$

The divergence at $x = x'$ is allowed for by taking the Cauchy principal value of the integral. It will be seen that $F_{\mathrm{Hi}}(x)$ is a linear functional of $f(x)$; in fact it is obtainable from $f(x)$ by convolution with $(-\pi x)^{-1}$. To emphasize this relationship, we may write

$$F_{\mathrm{Hi}} = \frac{-1}{\pi x} * f(x).$$

From the convolution theorem we can now say how the spectrum of $F_{\mathrm{Hi}}(x)$ is related to that of $f(x)$.

The Fourier transform of $(-\pi x)^{-1}$ is $i \operatorname{sgn} s$ (Fig. 13.9), which is equal to $+i$ for positive s and $-i$ for negative s; hence Hilbert transformation is equivalent to a curious kind of filtering, in which the amplitudes of the spectral components are left unchanged, but their phases are altered by $\pi/2$, positively or negatively according to the sign of s (Fig. 13.9).

Since the application of two Hilbert transformations in succession reverses the phases of all components, it follows that the result will be the negative of the original function. Hence

$$f(x) = -\left(\frac{-1}{\pi x}\right) * F_{\mathrm{Hi}}$$

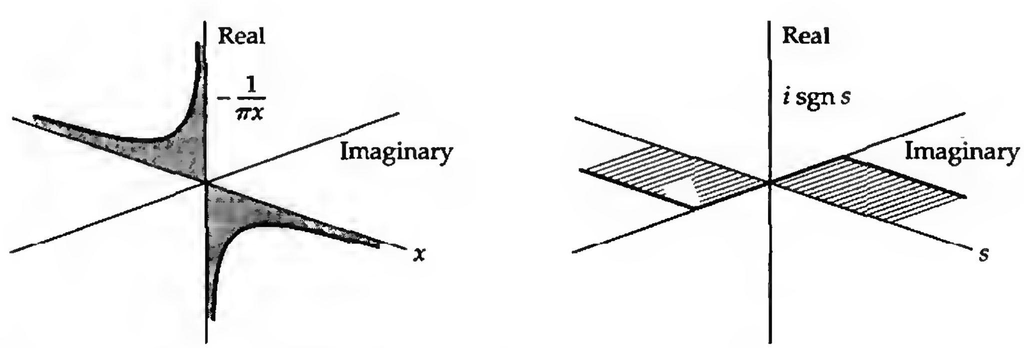


Fig. 13.9 **The kernel $(-\pi x)^{-1}$ and its Fourier transform i sgn s.**

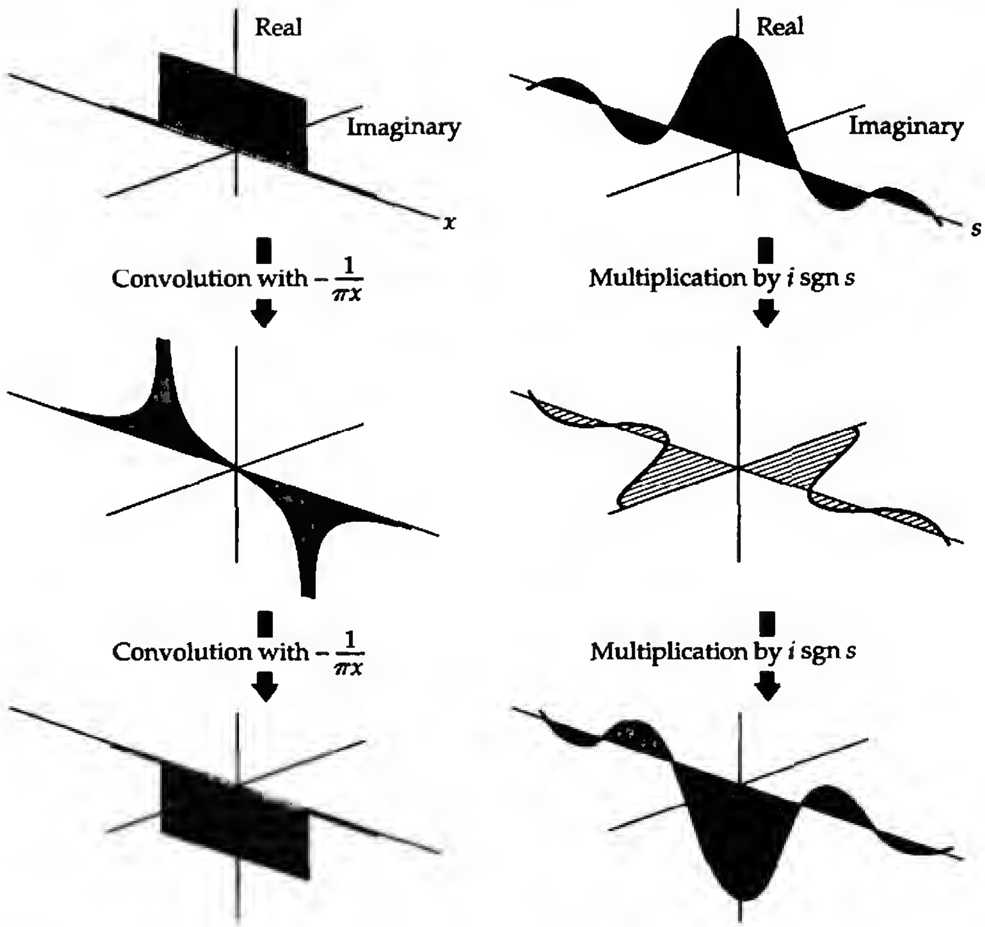


Fig. 13.10 **The left-hand column shows a function and the result of two successive Hilbert transformations, while the right-hand column shows the corresponding Fourier transforms.**

or
$$f(x) = -\frac{1}{\pi}\int_{-\infty}^{\infty} \frac{F_{\mathrm{Hi}}(x')\,dx'}{x' - x}.$$

Had the kernel been chosen as $[i\pi(x' - x)]^{-1}$ instead of $[\pi(x' - x)]^{-1}$, the transformation would have been strictly reciprocal, for then the effect would have been to multiply the spectrum by sgn s, and two such multiplications produce no net change. The custom is to sacrifice the symmetry which would be gained by this procedure in favor of the property that the Hilbert transform of a real function should also be a real function.

It will be noticed that all cosine components transform into negative sine components and that all sine components transform into cosines (Fig. 13.11). A consequence of this is that the Hilbert transforms of even functions are odd and those of odd functions even.

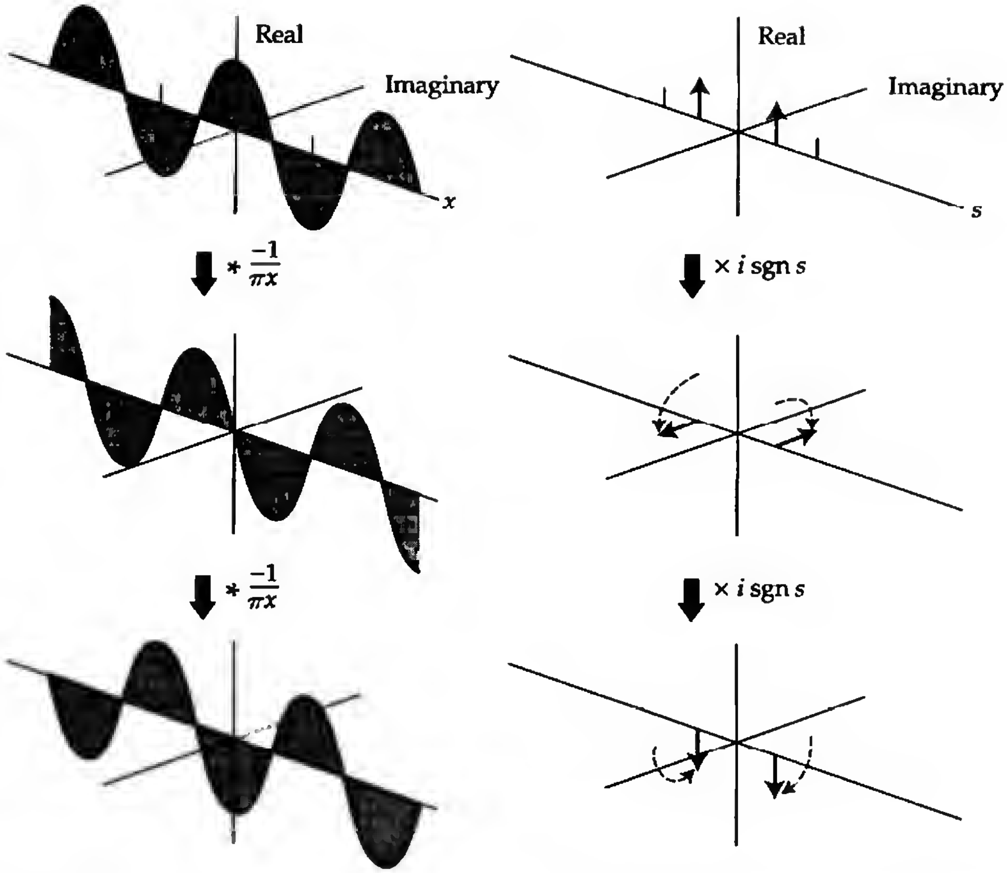


Fig. 13.11 **Explaining the Hilbert transform of cosine and sine functions (left) in terms of their spectra (right).**

The analytic signal. Consider a real function $f(t)$. With it we may associate a complex function

$$f(t) - iF_{\mathrm{Hi}}(t),$$

whose real part is $f(t)$. In signal analysis and optics, where the independent variable is time, this associated complex function is known as the analytic signal, and the Hilbert transform is referred to as the quadrature function of $f(t)$. As an example, the quadrature function of $\cos t$ is $-\sin t$ and the analytic signal corresponding to $\cos t$ is $\exp it$. One might say that the analytic signal bears the same relationship to $f(t)$ as $\exp it$ does to $\cos t$.

Just as phasors simplify manipulations in a-c theory, so the analytic signal is useful in some situations in which departure from absolutely monochromatic behavior prevents use of phasors. Modulated carriers and laser signals furnish examples. Figure 13.12 shows a function $f(t)$ that could be regarded as a modulated carrier; that is, it is a quasi-monochromatic pulse that slowly builds up and dies

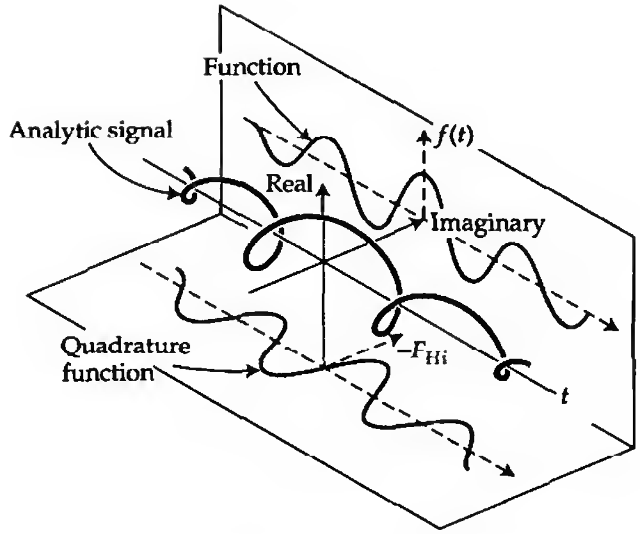


Fig. 13.12 **An amplitude modulated carrier $f(t)$, its quadrature function $F_{\mathrm{Hi}}(t)$, and the associated complex analytic signal.**

away. Its Hilbert transform is also shown, and together the two constitute the complex function of time represented by a helix that slowly dilates and contracts. The original signal $f(t)$ is the projection of this twisted curve on the plane defined by the time axis and the axis of reals (the vertical plane shown), and the quadrature function $F_{\mathrm{Hi}}(t)$ is the projection on the horizontal plane.

If a mean angular frequency $\bar{\omega}$ can be assigned, the analytic signal can be written

$$V(t)e^{i\bar{\omega}t},$$

where the complex coefficient $V(t)$, since it is a generalization of the phasor, can be described as a "time-varying phasor." In the case illustrated in Fig. 13.12, $V(t)$ is real and slowly rises and falls in value. If there were some frequency modulation, $V(t)$ would also rock backward and forward in phase.

From the explanation it is clear that the analytic signal contains no negative-frequency components; in fact it is obtainable from $f(t)$ by suppressing the negative frequencies. For example, noting that

$$\cos \omega t = \frac{e^{i\omega t} + e^{-i\omega t}}{2},$$

we obtain the analytic signal by suppressing the negative-frequency term exp $(-i\omega t)$. It is also necessary to double the result. To show this directly, let $f(t) \supset F(f)$ and let $\hat{f}(t)$ be derived by suppressing the negative frequencies and doubling:

$$\hat{f}(t) \supset 2H(f)F(f).$$

since $$H(f) \subset \tfrac{1}{2}\delta(t) + \frac{i}{2\pi t},$$

it follows that
$$\hat{f}(t) = 2\left[\tfrac{1}{2}\delta(t) + \frac{i}{2\pi t}\right] * f(t)$$
$$= f(t) - i\left(\frac{-1}{\pi t}\right) * f(t)$$
$$= f(t) - iF_{\mathrm{Hi}}(t).$$

Thus $\hat{f}(t)$ is indeed the analytic signal, as originally defined.

Instantaneous frequency and envelope. There is some question how one might define the envelope of a signal, since it only touches the signal waveform from time to time and might do anything in between. Similarly, the frequency of a quasi-monochromatic signal is not obviously defined from moment to moment. However, one does not hesitate to say what the amplitude of a sine wave is between peaks; furthermore, there is equal confidence about the moment-to-moment frequency as long as the maxima and minima, and the zero-crossing intervals remain steady. Since the analytic signal $V(t)\,exp\,i\bar{\omega}t$ is defined for continuous t we may say that, in the not strictly sinusoidal, or monochromatic, case $|V(t)|$ is the instantaneous amplitude, or envelope, and that the time rate of change of the phase of the analytic signal is the instantaneous frequency.

To compute instantaneous amplitude from a given $f(t)$ requires adoption of a mean angular frequency $\bar{\omega}$, a parameter that may be assigned with confidence to some waveforms. For example, with an AM radio signal you would take $\bar{\omega}$ to be the carrier frequency. But in general, $\bar{\omega}$ and $V(t)$ are not uniquely defined; consequently, when you see the term envelope used for $V(t)$ be prepared for surprises.

Causality. It is well known that effects never precede their causes, and so $I(t)$, the response of a physical system to an impulse applied at $t = 0$, must be zero for negative values of t:

$$I(t) = 0 \qquad t < 0.$$

Impulse responses satisfying this condition are said to be "causal." This condition is sometimes called the condition of "physical realizability," which is a bad term because it ignores the fact that a system may be impossible to construct for quite other reasons.[3]

Consider some real function $J(t)$. The most general causal impulse response has the form

$$I(t) = H(t)J(t).$$

The spectrum of $I(t)$, namely the transfer function $T(f)$, where

$$T(f) = \int_{-\infty}^{\infty} I(t)e^{-i2\pi ft}\,dt,$$

[3] In the literature the term "physical realizability" may include a second condition having nothing to do with causality, namely, that $I(t)$ should die away with time in such a way as to eliminate circuits, such as transmission-line segments, that do not have a finite number of degrees of freedom.

must have a special property corresponding to the causal character of $I(t)$, and from the previous discussion of the analytic signal this property must be that the real and imaginary parts of $T(f)$ are a Hilbert transform pair. Split $I(t)$ into its even and odd parts:

$$\begin{aligned} I(t) &= E(t) + O(t) \\ &= \tfrac{1}{2}[I(t) + I(-t)] + \tfrac{1}{2}[I(t) - I(-t)]. \end{aligned}$$

Now since $O(t) = \operatorname{sgn} tE(t)$, and $\operatorname{sgn} t \supset -i/\pi f$,

$$\begin{aligned} I(t) &= (1 + \operatorname{sgn} t)E(t) \\ &\supset G(f) + i\left(\frac{-1}{\pi f}\right) * G(f). \end{aligned}$$

Hence all causal transfer functions $T(f)$ are of the form

$$T(f) = G(f) + iB(f),$$

where $B(f)$ is the Hilbert transform of $G(f)$; that is,

$$B(f) = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{G(f')}{f' - f} df',$$

and

$$G(f) = -\frac{1}{\pi}\int_{-\infty}^{\infty} \frac{B(f')}{f' - f} df'.$$

Several Hilbert transforms are listed in Table 13.11.

COMPUTING THE HILBERT TRANSFORM

It would appear that the Hilbert transform of $f(x)$ could be computed as follows. First discretize at unit interval in x. Take the discrete Fourier transform, multiply by $i \operatorname{sgn} s$ (which is the Fourier transform of $-i/\pi x$), and invert the transform. It is simpler, however, to discretize $f(x)$ and convolve with a set of coefficients representing $-1/\pi x$. For example, if a semispan of $N = 5$ sample spacings was deemed sufficient the $2N + 1$ coefficients would be

$$\{.064\ \ .080\ \ .106\ \ .159\ \ .318\ \ 0\ \ -.318\ \ -.159\ \ -.106\ \ -.080\ \ -.064\}.$$

It might seem more satisfactory to restrict the dependence of each Hilbert transform value to the immediate vicinity than to involve the behavior of $f(x)$ far away. The reason for concern about distant places has to do with the coefficients that are discarded when a span of only 10 is adopted (or any other finite span), because the sum of the discarded coefficients is infinite (on the left) or minus infinity (on the right). By convolving with the finite set of coefficients we trust that the two infinities will cancel, but how can they after data samples have been multiplied in? A second concern is the discrete sampling interval, which must be kept

TABLE 13.11
Some Hilbert transforms

$f(x)$	$F_{\mathrm{Hi}}(x)$
const.	0
$\cos x$	$-\sin x$
$\sin x$	$\cos x$
$\cos^2 x$	$-\frac{1}{2}\sin 2x$
$\sin^2 x$	$\frac{1}{2}\sin 2x$
$\operatorname{cas} x$	$\operatorname{cas}(-x)$
e^{ix}	ie^{ix}
$x^{-1}\sin x$	$x^{-1}(\cos x - 1)$
$\Pi(x)$	$\pi^{-1}\ln \lvert (x-\frac{1}{2})/(x+\frac{1}{2})\rvert$
$\Lambda(x)$	$\pi^{-1}\{\ln \lvert (x-1)/(x+1)\rvert + \ln \lvert x^2/(x^2-1)\rvert\}$
$\Lambda(x)\operatorname{sgn} x$	$-\pi^{-1}\{(1-x)\ln \lvert x/(x+1)\rvert + 1\}$
$1/(1+x^2)$	$-x/(1+x^2)$
$\operatorname{sinc}' x$	$-\pi \operatorname{sinc} x - \frac{1}{2}\pi \operatorname{sinc}^2 \frac{1}{2}x$
$\delta(x)$	$-1/\pi x$
$\delta'(x)$	$1/\pi x^2$
$\delta''(x)$	$-2/\pi x^3$
$\lvert x\rvert^{-1/2}$	$\lvert x\rvert^{-\frac{1}{2}}$
$\lvert x\rvert^{-\frac{1}{2}}\operatorname{sgn} x >$	$-\lvert x\rvert^{-\frac{1}{2}}\operatorname{sgn} x$
$(1-x^2)\Pi(x/2)$	$-\pi^{-1}\{(1-x^2)\ln \lvert (x-1)/(x+1)\rvert - 2x\}$
$(1-x^2)^{\frac{1}{2}}\Pi(x/2)$	$-x + (x^2-1)^{\frac{1}{2}}[1-\Pi(x/2)]\operatorname{sgn} x$
$(1+x)/(1+x^2)$	$(1-x)/(1+x^2)$
$\text{п}(x)$	$x/\pi(\frac{1}{4}-x^2)$
$\text{ɪɪ}(x)$	$-\frac{1}{2}\pi(\frac{1}{4}-x^2)$
$\text{III}(x)$	$-\sum_{n=-\infty}^{\infty}\cot[\pi(x-n)]$

small enough to work against aliasing of signal content where $|s| > 0.5$. This concern might seem to be alleviated by the Fourier transform approach, but in fact it is not.

First note that the convolving function $h(x)$, represented as a string of impulses

$$h(x) = \sum_{n=1}^{N}\left[\frac{1}{\pi n}\delta(x+n) - \frac{1}{\pi n}\delta(x-n)\right]$$

has a Fourier transform

$$H(s) = \sum_{n=1}^{N} \frac{2i}{\pi n} \sin 2\pi n s$$

that can be compared with the desired transfer function $i \operatorname{sgn} s$ (Fig. 13.13). The phase is correct, $\pi/2$ for positive s and $-\pi/2$ for negative s, but the amplitude instead of having magnitude unity as desired (see the fine line), discriminates against both high and low frequencies, treating only a relatively narrow band peaking near $0.45N$ with approximately unity transfer function. Samples of $1/\pi x$ taken at $x = \pm\frac{1}{2}, 1\frac{1}{2}, \ldots$ instead of at eleven integer values of x lead to the same outcome.

Now the transfer function achieved by the Fourier method fails similarly, hugging the line from (0, 1) to (0.5, 0) more or less tightly according to the length of the data set, despite the fact that the multiplying factor $i \operatorname{sgn} s$ has unit amplitude for all s. The saw-tooth behavior can be traced to unavoidable aliasing due to under sampling.

One can obtain a satisfactory envelope for a narrow-band oscillatory function such as the sunspot number series (Bracewell, 1985) but in general it is recommended that the $N = 5$ convolution method be tried on a few sinusoidal signals in the frequency band occupied by the data. Drooping of the transfer function below unity can then be corrected empirically. Pulsations that appear on the envelope between the expected points of tangency to the data can be reduced by adjusting the discretizing interval. Further improvement can be sought by changing to the Fourier method and adjusting the Fourier transform by division by $|H(s)|$ over the band of interest.

Hilbert transformation can be performed numerically on any finite-duration signal but has some unpleasant attributes that are not obvious when the Hilbert transform is invoked in an analytic context.

Hilbert transformation may be implemented using the discrete Hartley transform as follows. Let the data $f(\tau)$ be specified for values of τ running from 0 to $N - 1$. Take the Hartley transform $H(\nu)$ and swap $N/2 - 1$ pairs of values to form

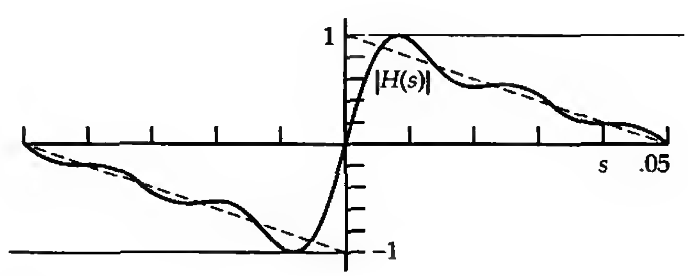
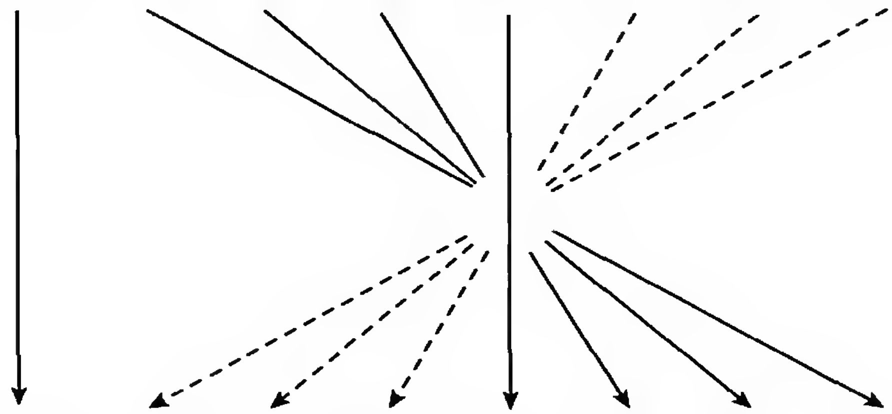


Fig. 13.13 **The transfer function with amplitude sgn *s* required for Hilbert transformation (fine line) compared with the transfer function actually achieved (curve) with the 11-element convolving sequence. The Fourier method applied using the same discretizing interval fails in a similar way, hugging the broken line.**

$H_2(\nu)$. In each swap, one of the elements changes sign as indicated. Thus

replace $H = \{H(0) \quad H(1) \quad H(2) \quad H(3) \quad H(4) \quad H(5) \quad H(6) \quad H(7)\}$



by $H_2 = \{H(0) \quad -H(7) \quad -H(6) \quad -H(5) \quad H(4) \quad H(3) \quad H(2) \quad H(1)\}$

Take the inverse Hartley transform of $f(t)$ to get the Hilbert transform (Pei and Jaw, 1989).

THE FRACTIONAL FOURIER TRANSFORM

As noted in Chapter 2, the repeated Fourier transformation $\mathcal{F}\mathcal{F}f(x)$ yields $f(-x)$ while four applications $\mathcal{F}\mathcal{F}\mathcal{F}\mathcal{F}f(x)$ comes full circle and yields the original $f(x)$. In two dimensions, $\mathcal{F}\mathcal{F}f(x,y)$ yields $f(-x,-y)$, which can be regarded as a rotation of $f(x,y)$ through an angle π, while $\mathcal{F}\mathcal{F}\mathcal{F}\mathcal{F}f(x,y)$, or $\mathcal{F}^4$ for short, introduces a further rotation π that returns to the original $f(x,y)$.

Letting the notation $\mathcal{F}^a$ represent a applications of the Fourier transform we now have interpretations $\mathcal{F}^a$ for $a = 1, 2$, and 4, and can readily generalize to any integer value of a such as 0 and 3; the case $\mathcal{F}^{-1}$ would be the inverse Fourier transformation, whose kernel is exp $i2\pi sx$. The Fourier transform of order a, where a is an integer, exhibits

consistency	$\mathcal{F}^a = \mathcal{F}$ when $a = 1$,
additivity	$\mathcal{F}^a\mathcal{F}^b = F^{a+b}$,
commutativity	$\mathcal{F}^a\mathcal{F}^b = F^bF^a$,
and linearity	$\mathcal{F}^a(f + g) = \mathcal{F}^a f + \mathcal{F}^a g$.

Is there an algebraic generalization that retains these group-theoretical properties when a is not restricted to being an integer? For example, can meaning be given to a half-order Fourier transform $\mathcal{F}^{\frac{1}{2}}$? It would be required to possess the above properties; in particular, half-order transformation applied twice in succession should exhibit additivity and thus yield the ordinary Fourier transform.

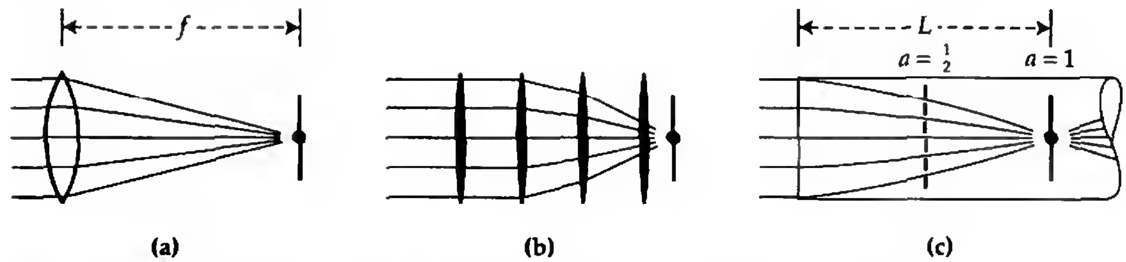


Fig. 13.14 **Three focusing arrangements that produce Fourier transforms in the planes marked by •. (a) A lens of focal length *f*. (b) A spaced set of thin lenses. (c) An optical fiber with radially graded refractive index.**

When the FFT is computed as described in Chapter 11, the data set of N values is replaced in successive stages until a final set is reached that is the desired transform. The intermediate stages can be viewed as partial transformations, and the middle one might be a half-order candidate. To follow up this idea one would have to deal with the fact that there is a middle stage only for $N = 8, 32, \ldots$ and with similar restrictions for values of a other than one-half. Permutation would have to be distributed over the stages too, instead of being concentrated at the beginning or end, and the resulting flow diagrams for each stage would have to be the same.

This discussion nevertheless offers a concept for developing a fast algorithm for computing purposes. A physical approach is offered by a two-dimensional example from diffraction theory, where a field distribution $f(x,y)$ in one plane can excite in a parallel plane, by Fraunhofer diffraction, a field distribution that is the two-dimensional transform of $f(x,y)$. Therefore the field distribution in an intermediate plane is a candidate for a fractional transform with $0 < a < 1$. Radiation diffracted by an aperture in the first plane spreads out in space, ultimately approaching the Fourier transform on the plane at infinity, a configuration that can be compacted by insertion of a simple lens of focal length f behind the (x,y)-plane, whereupon the transform will appear nearby in the focal plane of the lens at $z = f$. A difficulty arises that is reminiscent of permutation which, in the computing case, is normally concentrated at a single stage. In the lens case, it is the focusing that is concentrated at one place.

However, focusing in practice can be distributed (Fig. 13.14); one way is to replace the lens by several thin lenses distributed between the function plane and the transform plane. In addition continuous focusing is already available in certain types of existing optical fiber, any slice of which is a disc that acts just like a thin lens. A lens achieves focusing by being thicker on axis and thus delaying the axial ray relative to the outer rays as required for all to arrive at the focus in the same phase; a disc sliced from a fiber or rod can delay the axial ray by having a greater refractive index on axis.

Let the refractive index n fall off quadratically from a central value n_0 on the axis (just as lens thickness varies quadratically) according to the formula $n = n_0(1 - r^2/h^2)$. Such a dielectric is known as a graded refractive index medium.

A field distribution $f(x,y,0)$ on the entry plane $z = 0$ gives rise to the Fourier transform (corresponding to $a = 1$) in the plane $z = L$, where $L = \pi h/2\sqrt{2n_0}$. The half-order transform ($a = \frac{1}{2}$) appears half way and other fractional orders appear at distances aL from the aperture. The detailed theory of wave propagation in a graded medium is the background for the definition adopted below for the fractional-order Fourier transform.

Reverting to a one-dimensional independent variable, as exemplified by time t, let $f(t)$ have a Fourier transform $F(f)$, and define its fractional Fourier transform of order a, where a is a fraction between 0 and 1, by

$$\mathcal{F}^a f(t) = F_a(f) = \frac{e^{i(\frac{1}{4}\pi - \frac{1}{2}\phi)}}{\sqrt{\sin\phi}} \int_{-\infty}^{\infty} f(t) \exp\left[-i2\pi \frac{ft - \frac{1}{2}(t^2 + 4\pi^2 f^2)\cos\phi}{\sin\phi}\right] dt,$$

where $\phi = \frac{1}{2}\pi a.$ This definition is derived from Namias (1980) by elimination of a factor $\sqrt{2\pi}$ to conform with system 1 of Chapter 2. Testing for consistency by putting $a = 1$ we find $F_1(f) = \int_{-\infty}^{\infty} f(t)\exp(-i2\pi ft)dt$, which is the same as $F(\,)$. For $a \to 0$ a painstaking limiting process (McBride and Kerr, 1987) confirms that the defining expression approaches $f(\,)$, the original function. Use of frequency f, whose unit is one cycle per unit of t, as the independent variable for the transform lends some familiarity to the expressions. An unfamiliar feature is that the fractional transform glides continuously from the time domain to the frequency domain; this peculiarity is glossed over if the independent variable is taken to be dimensionless. Some theorems and examples of fractional transforms are as follows.

Shift theorem. If $f(t)$ has fractional Fourier transform $F_a(f)$ then

$$\mathcal{F}^a f(t - T) = e^{-iT\sin\phi(2\pi f - \frac{1}{2}T\cos\phi)} F_a(f - T\cos\phi).$$

Derivative theorems.

$$\mathcal{F}^a f'(t) = \left(i2\pi f + \cos\phi \frac{d}{df}\right) F_a(f),$$

$$\mathcal{F}^a f''(t) = \left(i2\pi f + \cos\phi \frac{d}{df}\right)^2 F_a(f),$$

$$\mathcal{F}^a t f(t) = -\left(2\pi f \cos\phi + i \sin\phi \frac{d}{2\pi df}\right) F_a(f),$$

$$\mathcal{F}^a t^2(t) = \left(2\pi f \cos\phi + i \sin\phi \frac{d}{2\pi df}\right)^2 F_a(f),$$

Fractional convolution theorem.

$$f(t) *^a g(t) = e^{-ibt^2} \int_{-\infty}^{\infty} f(\tau) e^{ib\tau^2} g(t - \tau) e^{ib(t-\tau)^2} d\tau,$$

where $b = \frac{1}{2}\cot(\frac{1}{2}\pi a)$.

Examples of transforms.

$$\mathcal{F}^a \delta(t) = \frac{e^{i(\frac{1}{4}\pi - \frac{1}{2}\phi)}}{\sqrt{\sin\phi}} e^{i4\pi^3 f^2 \cot\phi}.$$

$$\mathcal{F}^a\delta(t-T)=\frac{e^{i(\frac{1}{4}\pi-\frac{1}{2}\phi)}}{\sqrt{\sin\phi}}e^{-i\frac{2\pi fT-\pi(4\pi^2f^2+T^2)\cos\phi}{\sin\phi}}.$$

$$\mathcal{F}^a[\delta(t+T)+\delta(t-T)]=\frac{e^{i(\frac{1}{4}\pi-\frac{1}{2}\phi)}}{\sqrt{\sin\phi}}e^{i\pi(4\pi^2f^2+T^2)\cot\phi}\,2\cos\left(\frac{2\pi fT}{\sin\phi}\right).$$

$$\mathcal{F}^aH_n(\sqrt{2\pi}t)e^{-\pi t^2}=e^{in\alpha}H_n(\sqrt{2\pi}s)e^{-\pi f^2},$$

where the $H_n(x)$ are Hermite polynomials (see Problem 8.25 for listing), and the orthogonal functions $H_n(\sqrt{2\pi}x)e^{-\pi t^2}$ are eigenfunctions of the fractional Fourier transform.

Applications. Applications to solving Schrödinger's and other second order differential equations in one or more dimensions are illustrated by Namias (1980). Applications in optics, especially to optical fibers (Mendelovic and Ozaktas, 1993; Ozaktas and Mendelovic, 1993a) and to optical information processing systems (Ozaktas and Mendelovic, 1993b) have been reported and followed up. Lohmann (1995) has related the fractional Fourier transform to rotation of the Wigner distribution with implications for time-frequency analysis of signals (Mendlovic et al., 1996). Sheppard (1998) has derived a relationship between free-space diffraction and the fractional Fourier transform, while Kutay and Ozaktas (1998) have demonstrated restoration of optical images.

Noting varying disadvantages of working with complex integrals Mendlovic et al. (1995) have demonstrated a fractional optical transform that is real, starting from a technique of Bracewell et al. (1985), for the two-dimensional optical Hartley transform. If $V(t)$ has one-dimensional Hartley transform $H(f)=\int_{-\infty}^{\infty}\text{cas}\,2\pi ft\,dt$ then for the fractional Hartley transform of order a one has

$$H_a(f)=\frac{1}{\sqrt{\sin\phi}}\int_{-\infty}^{\infty}f(t)\text{cas}\left[(\tfrac{1}{4}\pi-\tfrac{1}{2}\phi)+2\pi\frac{ft-\frac{1}{2}(t^2+4\pi^2f^2)\cos\phi}{\sin\phi}\right]dt.$$

Future applications may be expected in the area of lens system design. Other branches of physics where situations analogous to optical guiding occur include ion beams, where magnetic focusing is utilized (Pierce, 1954), and magnetospheric physics, where 5 to 15 kHz radio waves from lightning escape into ducts that guide the radiation to the opposite hemisphere (Helliwell, 1965). It is striking that the quantum mechanical harmonic oscillator wave functions are the same as the natural modes of a graded-index optical fiber and of a laser cavity and at the same time are the eigenfunctions of the fractional Fourier transform operator. The fractional Fourier transform offers a new window on old topics.

For an unrelated confusingly-named discrete transform based on fractional roots of unity rather than the conventional Nth roots of unity $\exp(-i2\pi/N)$, see Bailey and Swartztrauber (1990).

BIBLIOGRAPHY

Ahmed, A., T. Natarajan and K. Rao: "Discrete Cosine Transform," *IEEE Trans. Comp.*, vol. C-23, pp. 90–03, 1974.

Bailey, D. H., and P. N. Swartztrauber: "The Fractional Fourier Transform and Applications," *SIAM Review*, vol. 33, 389–404, 1990.

Bracewell, R. N.: "Sunspot Series Envelope and Phase," *Austral. J. Physics*, vol. 38, pp. 1009–1025, 1985.

Bracewell, R. N.: "Affine Theorem for the Hartley Transform of an Image," *Proc. IEEE*, vol. 82, pp. 381–387, 1994.

Bracewell, R. N.: "Two-dimensional Imaging," Prentice-Hall, Englewood Cliffs, New Jersey, 1995.

Bracewell, R. N., and A. C. Riddle: "Inversion of Fan Beam Scans in Radio Astronomy," *Astrophys. J.*, vol. 150, pp. 427–434, 1967.

Bracewell, R. N., K.-Y. Chang, A. K. Jha, and Y.-H. Wang: *Electronics Letters*, vol. 29, p. 304, 1993.

Bracewell, R. N., H. Bartelt, A. W. Lohmann, and N. Streibl: "Optical Synthesis of the Hartley Transform," *Appl. Opt.*, vol. 24, pp. 1401–1402, 1985.

Deans, S. R.: "The Radon Transform and Some of Its Applications," Wiley, New York, 1983.

Erdélyi, A. ed.: "Tables of Integral Transforms," McGraw-Hill, New York, 1954.

Helliwell, R. A.: "Whistlers and Related Ionospheric Phenomena," Stanford University Press, Stanford, Calif., 1965.

Lohmann, A. W.: "Image Rotation, Wigner Rotation, and the Fractional Fourier Transform," *J. Opt. Soc. Am. A*, vol. 10, pp. 2181–2186, 1993.

Lohmann, A. W.: "A Fake Zoom Lens for Fractional Fourier Experiment," *Opt. Comm.*, vol. 115, pp. 437–443, 1996.

Lohmann, A. W., D. Mendlovic, and Z. Zalevsky: "Fractional Transformations in Optics," *Prog. in Optics*, vol. 38, pp. 263–342, 1998.

McBride, A. C., and F. H. Kerr: "On Namias's Fractional Fourier Transform," *IMA J. Appl. Math.*, vol. 39, pp. 159–175, 1987.

Mendlovic, D., and H. M. Ozaktas: "Fractional Fourier Transformations and Their Optical Implementation, I," *J. Opt. Soc. Am. A*, vol. 10, pp. 1875–1881, 1993.

Mendlovic, D., Z. Zalevsky, N. Konforti, R. G. Dorsch, and A. W. Lohmann: "Incoherent Fractional Fourier Transform and Its Optical Implementation," *Appl. Opt.*, vol. 34, pp. 7615–7620, 1995.

Mendlovic, D., H. D. Ozaktas, and A. W. Lohmann: "Graded-index Fibers, Wigner-distribution Functions and the Fractional Fourier Transform," *Appl. Opt.*, vol. 33, pp. 6188–6193, 1994.

Mendlovic, D., H. D. Ozaktas, and A. W. Lohmann: "Graded-index Fibers, Wigner-distribution Functions, and the Fractional Fourier Transform," *Appl. Comm.*, vol. 33, pp. 6188–6193, 1996.

Namias, V.: "The Fractional Order Fourier Transform and Its Application to Quantum Mechanics," *J. Inst. Maths Apllics.*, vol. 25, pp. 241–265, 1980.

Ozaktas, H. M., and D. Mendlovic: "Fractional Fourier Transformations and Their Optical Implementation, II," *J. Opt. Soc. Am. A*, vol. 10, pp. 2522–2531, 1993a.

Ozaktas, H. M., and D. Mendlovic: "Fourier Transforms of Fractional Order and Their Optical Interpretation," *Opt. Comm.*, vol. 101, pp. 163–169, 1993b.

Ozaktas, H. M., B. Barshan, D. Mendlovic, and L. Onural: "Convolution, Filtering, and Multiplexing in Fractional Fourier Domains and Their Relation to Chirp and Wavelet Transforms," *J. Opt. Soc. Am. A*, vol. 11, pp. 547–559, 1994.

Pei, S-C, and S-B Jaw: "Discrete Hilbert Transform by FHT," *IEEE Trans. Circuits Systems*, vol. 36, pp. 1251–1252, 1989.

Pierce, J. R.: "Theory and Design of Electron Beams," Van Nostrand, New York, 1954.

Sheppard, C. J.: "Free-space Diffraction and the Fractional Fourier Transform," *J. Mod. Opt.*, vol. 45, pp. 2097–2103, 1998.

Wang, S., and N. George: "Fresnel Zone Transforms in Spatially Incoherent Illumination," *Appl. Opt.*, vol. 24, pp. 842–850, 1985.

PROBLEMS

1. Demonstrate the following theorems relating to a function $f(x)$ and its Hilbert transform $F_{\mathrm{Hi}}(x)$.

(*a*) $$f(ax) \supset F_{\mathrm{Hi}}(ax) \qquad \text{(similarity)}$$

(*b*) $$f(x) + g(x) \supset F_{\mathrm{Hi}}(x) + G_{\mathrm{Hi}}(x) \qquad \text{(addition)}$$

(*c*) $$f(x-a) \supset F_{\mathrm{Hi}}(x-a) \qquad \text{(shift)}$$

(*d*) $$\int_{-\infty}^{\infty} f(x)f^*(x)\,dx = \int_{-\infty}^{\infty} F_{\mathrm{Hi}}(x)F^*_{\mathrm{Hi}}(x)\,dx \qquad \text{(power)}$$

(*e*) $$\int_{-\infty}^{\infty} f^*(x)f(x-u)\,dx = \int_{-\infty}^{\infty} F^*_{\mathrm{Hi}}(x)F_{\mathrm{Hi}}(x-u)\,dx$$

(autocorrelation)

(*f*) $$f * g \supset -F_{\mathrm{Hi}} * G_{\mathrm{Hi}} \qquad \text{(convolution)}$$

2. **Hilbert transform of convolution.** Let $\mathcal{H}f$ represent the Hilbert transform of $f(x)$. Show that the Hilbert transform of a convolution can be expressed as follows:

$$\mathcal{H}(f * g) = \mathcal{H}f * g = f * \mathcal{H}g.$$

3. **Autocorrelation.** Explain why a function and its Hilbert transform have the same autocorrelation function.

4. **Uniqueness.** It is said that the Hilbert transformation is not unique because the transform of unity is zero. Consequently, the inverse transform of zero would be any constant, and therefore the Hilbert transform of any function would be uncertain to the extent of an additive constant. Examine this argument critically and make an authoritative report on the question of uniqueness.

5. What are the analytic signals corresponding to the following waveforms: sinc t, $\exp[-(t-t_0)^2]\cos\omega t$, $(1 + M\cos\Omega t)\cos\omega t$? Make three-dimensional sketches showing the analytic signal and its two projections.

6. **Creating Hilbert pairs.** Generate Hilbert transform pairs by separating the real and imaginary parts of the Fourier transforms of the following physically realizable impulse-response functions: $\exp(-t)H(t)$, $t\exp(-t)H(t)$, $\Lambda(t-1)$, $\exp(-t)\cos\omega t\,H(t)$.

7. Generate Hilbert transform pairs by taking the Fourier transform of the odd and even parts of the following physically realizable impulse-response functions:

$$\Lambda(t)H(t),\ \Pi(t-\tfrac{1}{2}).$$

8. A real function of t is zero for $t < 1$. Prove that the real and imaginary parts of its Fourier transform form a Hilbert transform pair.

9. **Precocious causality.** What can be said about the Fourier transform of a real function of t that is zero for $t < -1$?

10. Let the transfer function $T(f)$ of a filter be expressed as

$$T(f) = e^{\Theta(f)},$$

where $\Theta(f)$ is the complex electrical length. Show that $\Theta(f)$ is hermitian.

11. In the preceding problem, let

$$\Theta(f) = \alpha(f) + i\beta(f),$$

where $\alpha(f)$ is the gain and $\beta(f)$ the phase change of the filter. Thus if

$$T(f) = G(f) + iB(f),$$

then

$$\alpha(f) = \log (G^2 + B^2)^{\frac{1}{2}} = \log |T|$$

and

$$\beta(f) = \tan^{-1}\frac{B}{G} = \text{pha}\, T.$$

It has been shown that $\Theta(f)$ is hermitian, that is, that $\Theta(-f) = \Theta^*(f)$, or that $\alpha(f)$ is even and $\beta(f)$ is odd. To prove this, was it only necessary to assume that $T(f)$ was hermitian [$I(t)$ real]? If so, what further property must $\alpha(f)$ and $\beta(f)$ exhibit if $T(f)$ is, in addition, causal?

12. **Causal filter.** If $G(f) + iB(f)$ is the transfer function of a causal filter, show that

$$B(f) = -\frac{2f}{\pi}\int_0^\infty \frac{G(u)}{f^2 - u^2}\,du$$

and

$$G(f) = \frac{2}{\pi}\int_0^\infty \frac{uB(u)}{f^2 - u^2}\,du.$$

13. What can be said about the Hilbert transform of a function that is hermitian?

14. **Band-limited functions.** Show that the Hilbert transform of a band-limited function is band-limited.

15. A function $f(t)$ is band-limited. It is said that from a knowledge of $f(t)H(-t)$ only, the whole of $f(t)$ can be deduced. Examine this statement.

16. Show that the Hankel transformation is equivalent to an Abel transformation followed by a one-dimensional Fourier transformation, or that

$$2\pi\int_0^\infty dr\, J_0(2\pi\xi r)r \int_{-\infty}^{\infty} ds\, e^{i2\pi rs} \int_s^\infty dx\, 2xf(x)(x^2 - s^2)^{-\frac{1}{2}} = f(\xi).$$

17. z transform. Let the impulse train

$$\sum_{0}^{\infty} f(n)\delta(t-n)$$

possess a Laplace transform where $\alpha < \operatorname{Re} p < \beta$. Show that the z transform $F(z)$ of the sequence $f(n)$ exists in an annulus of the z plane where $e^{-\beta} < |z| < e^{-\alpha}$.

18. Establish the following Hankel transform pairs.

$J_0(2\pi a r)$	$(2\pi a)^{-1}\,\delta(q-a)$
$J_0^2(2\pi a r)$	$\pi^{-2}q^{-1}(\pi^{-2}-q^2)^{-\frac{1}{2}}\Pi(\pi q/2)$
$\operatorname{jinc} r \equiv (2r)^{-1}J_1(\pi r)$	$\Pi(q)$
$\operatorname{jinc}^2 r$	$\frac{1}{2}[\cos^{-1} q - q(1-q^2)^{\frac{1}{2}}]\Pi(q/2)$
$r^{-n}J_n(r)$	$2\pi(1-4\pi^2q^2)^{n-1}\Pi(\pi q)/2^{n-1}(n-1)!$
$\exp(ir^2)$	$i\pi \exp(-i\pi^2 q^2)$
$\operatorname{sinc}^2 ar$	$(\pi a^2)^{-1}\cosh^{-1}(a/q)$

19. Shear theorem. If $f(x,y)$ has F.T. $F(u,v)$, show that $f(x,y-\alpha x)$ has F.T. $F(u+\alpha v, v)$.

20. Instantaneous musical pitch. A sound wave producing an air-pressure variation at the ear of $0.01 \cos[2\pi(500t + 50t^2)]$ newtons/meter^2 could be described as a moderately loud pure tone (54 decibels above reference level of 0.1 newton/meter^2) with a rising pitch. At $t = 0$ the frequency would be 500 hertz and rising at a rate of 100 hertz per second and after 2 or 3 minutes the sound would fade out as it rose through the limit of audibility.

(*a*) How long would it take for the pitch to rise 1 octave (1) starting from 500 hertz; and (2) starting from 6000 hertz?

(*b*) Calculate the rate of change of frequency in semitones per second.

(*c*) Construct a waveform that would be perceived by the ear as a pure tone having a uniformly rising pitch. Arrange that the tone rises through 500 hertz at a rate of 100 hertz per second.

21. Two-dimensional diffraction. A plowed field of 300 hectares is three times longer than it is wide and the furrows are three to the meter. The long axis of the field is oriented 20° east of north. On a (u,v) plane, with the u axis running east, sketch and dimension the principal features of the two-dimensional Fourier transform of $h(x,y)$, the height of the surface of the field. Conceive of a situation where the Fourier transform of a plowed field might arise.

22. Two-dimensional transform. An antenna in the form of a diamond-shaped or rhombic aperture has a long diagonal N wavelengths long and a short diagonal of n wavelengths. If N and n were equal, the aperture would be square and we know that the angular spectrum would be sinc u sinc v, and there would be null lines in directions $u = \pm 1, v = \pm 1$, etc. The following argument was presented. "The null lines are connected with the four directions of symmetry of a square aperture, two parallel to the sides and two parallel to the diagonals. But the rhombic aperture is symmetrical about its diagonals only. This leaves a lattice of discrete directions on which the angular spec-

trum is zero, but the null *lines* are destroyed no matter how little the corner angles depart from 90 degrees." Is that right?

23. X-ray diffraction. The atoms in a certain plane of a crystal lie on a square lattice of spacing 0.54 nanometer and give rise to a square pattern of diffraction spots when illuminated by an x-ray beam, as would be expected since the two-dimensional Fourier transform of ${}^2\text{III}(x,y)$ is ${}^2\text{III}(u,v)$. The crystal is now traversed by a microwave acoustic wave that places the plane mentioned in periodic shear. Opinions given by colleagues include the following.

(*a*) The diffraction spots will be displaced and the direction of displacement will depend on the wave direction in the plane.
(*b*) The spots will be widened in one direction only.
(*c*) The spots will be enlarged.
(*d*) The spot structure may be destroyed because the crystal strain could be large enough to prevent the constructive interference of the extremely short x-ray wavelengths on which the spots depend.
(*e*) Any effect would be so small compared with the spot size as to be undetectable.

Comment on these opinions.

24. Three-dimensional diffraction. X-rays incident on a crystal are scattered by the orbital electrons. Therefore, it is in principle possible by observing the scattered intensity in all directions to deduce the electron density in the crystal and then possibly to reason out the locations of the various atomic nuclei, even though the nuclei themselves are not probed. Let the electron density in a spherically symmetrical case be $\rho(r)$. Show that the intensity of scattering $K(s)$ is a function of one variable only and that $\rho(r)$ may be calculated from

$$\rho(r) = 4\pi \int_0^\infty K(s) \operatorname{sinc} 2rs \, ds.$$

What is the physical meaning of s and how could a crystalline substance in any way be regarded as spherically symmetrical?

25. Three-dimensional convolution. Consider a ball function $\Pi(\frac{r}{2})$, that is, a function of x, y, and z that is equal to unity inside a sphere of unit radius and zero outside. Show that its three-dimensional autocorrelation function (or self-convolution) is

$$\Pi\left(\frac{r}{2}\right) *** \Pi\left(\frac{r}{2}\right) = \left(\frac{2\pi}{3}\right)\left(2 - \frac{3r}{2} + \frac{r^2}{8}\right)\Pi\left(\frac{r}{4}\right).$$

26. Two-dimensional impulse. State the nature of the following impulse symbols in two dimensions by giving (*a*) the locus where the impulse is located and (*b*) the linear density at each point of the locus: $\delta(x + y)$, $\delta(xy)$, $\delta(\sin\theta)$, $\delta(x^2 + y^2 - 1)$, $\delta(x^2 + y^2)$.

27. Derivative theorems for Hankel transform. Show that

$$(rf)' \supset -(qF)'$$

and that $f' \supset -[q\mathcal{H}\{r^{-1}f\}]'$.

28. Derivative theorem for Hankel transform. Show that

$$rf'(r) \supset -q^{-1}\frac{d}{dq}[q^2F(q)].$$

29. **Hankel transform theorem.** Show that

$$f(r) = \mathcal{H}\left\{q^{-1}\frac{d}{dq}\mathcal{H}\left\{r^{-1}\frac{d}{dr}f(r)\right\}\right\}.$$

30. **Henkel transform example.** Establish that the Hankel transform of $r^2 \exp(-\pi r^2)$ is $(\pi^{-1} - q^2)\exp(-\pi q^2)$.

31. **Hankel transform.** Show that

$$\int_0^\infty J_1(x)J_0(ax)\,dx = H(1 - a^2).$$

32. **Hankel transform example.** Verify that $(4\pi r^2)^{-1}J_2(\pi r)$ has Hankel transform $(\frac{1}{4} - q^2)\Pi(q)$.

33. **Cauchy principal value.** We often use the phrase "area under the curve $f(x)$" to mean the integral from $-\infty$ to ∞. Intuitively, from experience with areas, one might expect that the area under $f(x)$ is the same as the area under $f(x + 1)$. Can you prove that

$$\int_{-\infty}^{\infty} \operatorname{sgn} x\,dx = \int_{-\infty}^{\infty} \operatorname{sgn}(x + 1)\,dx?$$

34. **Radial sampling under circular symmetry.** The light from a star is received at two points spaced a certain distance q apart and the complex correlation between the two optical waveforms is determined. It can be shown that this complex number is a value of the Hankel transform $B(q)$ of the brightness distribution $b(r)$ over the stellar disk (assuming that the brightness distribution has circular symmetry). (If r is measured in radians, q will be measured in wavelengths.) Since the star is of finite extent, it suffices to sample the transform at regularly spaced distances. Show how to determine $b(r)$ from values of $B(q)$ determined at $q = 0, a, 2a, \ldots$.

35. **Abel transform.** Let $f(\cdot)$ be subjected to two Abel transformations in succession. Show that the resulting function $f_{AA}(x)$ is equal to the volume under $f(\cdot)$ outside radius x, that is, $f_{AA}(x) = 2\pi\int_x^\infty rf(r)\,dr$. (This problem was supplied by S. J. Wernecke.)

36. **Two-dimensional autocorrelation.** Let $f(r)$ have Abel transform $f_A(x)$. If we take the two-dimensional autocorrelation of $f(r)$, we get another circularly symmetrical function. Show that the Abel transform of the two-dimensional autocorrelation is the one-dimensional autocorrelation of the Abel transform $f_A(x)$; that is, $f(r) ** f(r)$ has Abel transform $F_A(x) * f_A(x)$.

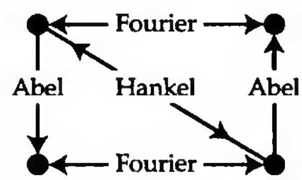
37. **Abel-Fourier-Hankel cycle of transforms.** Functions can be spatially arranged in groups of four to exhibit the Abel-Fourier-Hankel cycle of transforms (R. N. Bracewell, *Austral. J. Phys.*, vol. 9, p. 198, 1956, and Problem 13.16). Thus the relationships

jinc r	has Abel transform sinc x
sinc x	has Fourier transform $\Pi(q)$
$\Pi(q)$	has Hankel transform jinc r
$\Pi(q)$	has Abel transform $(1 - 4u^2)^{\frac{1}{2}}\Pi(u)$

$(1 - 2u^2)^{\frac{1}{2}}\Pi(u)$ has Fourier transform jinc r,

where jinc $r = (2r)^{-1}J_1(\pi r)$, are all compactly summarized by grouping the four functions as in the box.

jinc r	$(1 - 4u^2)^{\frac{1}{2}}\Pi(u)$
sinc x	$\Pi(q)$



The diagram on the right is the key to the transforms implied by the spatial relationship. Verify the following important groups.

sinc r	$\Pi(u)$
$J_0(\pi x)$	$\pi^{-1}(\frac{1}{4} - q^2)^{-\frac{1}{2}}\Pi(q)$

$\delta(r - a)$	$2 \cos 2\pi au$
$2a(a^2 - x^2)^{-\frac{1}{2}}\Pi\left(\dfrac{x}{2a}\right)$	$2\pi a J_0(2\pi aq)$

$M(r)$	$(1 - u^2)^{\frac{1}{2}} - u^2 \cosh^{-1} u^{-1}$
$\text{sinc}^2\, x$	$\Lambda(q)$

$e^{-\pi r^2}$	$e^{-\pi u^2}$
$e^{-\pi x^2}$	$e^{-\pi q^2}$

38. Verify the composite similarity theorem for the Fourier-Abel-Hankel cycle of transforms, for $a > 0$:

If

$f(r)$	$F(u)$
$g(x)$	$G(q)$

then

$af(ar)$	$F\left(\dfrac{u}{a}\right)$
$g(ax)$	$a^{-1}G\left(\dfrac{q}{a}\right)$

39. Noncircularly symmetrical Fourier transforms. Establish the following two-dimensional Fourier transform pairs.

$f(x,y)$	$F(u,v)$
$\delta(r - a)e^{in\theta}$	$(-i)^n 2\pi a e^{in\theta} J_n(2\pi aq)$
$\delta(r - a) \cos n\theta$	$(-i)^n 2\pi a \cos n\phi\, J_n(2\pi aq)$
$\delta(r - a) \sin n\theta$	$(-i)^n 2\pi a \sin n\phi\, J_n(2\pi aq)$
$\delta(r - a) \cos \theta$	$-i2\pi a \cos \phi\, J_1(2\pi aq)$
$\delta(r - a) \sin \theta$	$-i2\pi a \sin \phi\, J_1(2\pi aq)$
$\Pi\left(\dfrac{r}{2a}\right) \cos 4\theta$	$(2\pi q^2)^{-1} \cos 4\theta[4 + 8J_0(2\pi aq) + (2\pi aq - 12/\pi aq)J_1(2\pi aq)]$

40. Three-dimensional Fourier transforms. Establish the following three-dimensional Fourier transform pairs.

$k(r)$	$K(s)$
$\exp(-ar)$	$8\pi a[a^2 + (2\pi s)^2]^{-2}$
$r^{-1}\exp(-ar)$	$4\pi[a^2 + (2\pi s)^2]^{-1}$
$r^{-2}\exp(-ar)$	$2s^{-1}\tan^{-1}\left(\dfrac{2\pi s}{a}\right)$
$r^{-\frac{1}{2}}$	$(4\pi)^{-1}s^{-5/2}$
r^{-1}	$\pi^{-1}s^{-2}$
$r^{-3/2}$	$s^{-3/2}$
r^{-2}	πs^{-1}
$\exp(-\pi r^2)$	$\exp(-\pi s^2)$
$\operatorname{sinc}^2 r$	$(2\pi s)^{-1}\Pi\left(\dfrac{s}{4\pi}\right)$
$\operatorname{jinc} r \equiv (2r)^{-1}J_1(\pi r)$	$2\pi^{-1}(1-4s^2)^{-1/2}\Pi(s)$

41. Fourier transform in n dimensions. A function $f(r)$ in n-dimensional space is a function only of distance r from the origin, and its n-dimensional Fourier transform $F(q)$ is a function of only one variable q. For $n = 1, 2$, and 3 verify that the general formula

$$F(q) = 2\pi q^{(2-n)/2}\int_0^\infty r^{n/2}J_{(n-2)/2}(2\pi qr)f(r)\,dr$$

can claim to represent the n-dimensional Fourier transform.

42. Cosine transform. A function $f(x)$ is defined from 0 to ∞ by

$$f(x) = \begin{cases} 1 - 2x^2, & 0 < x < \frac{1}{2} \\ 0, & x > \frac{1}{2} \end{cases}$$

Find its cosine transform $F_c(s) = 2\int_0^\infty \cos 2\pi sx dx$. ▷

43. A business transform. A stock market analyst is accustomed to take daily stock prices $f(\tau)$ and to compute a transform

$$F_c(\nu) = \frac{1}{N}\sum_{\tau=-N+1}^{0} f(\tau)\cos(2\pi\nu\tau/N), \qquad 0 \leqslant N - 1$$

with a view to discerning seasonal trends that are combined with other information to advise clients. Usually N is taken as 365 and the transformation is repeated monthly. A new idea requires modification of $F_c(\nu)$, followed by a reverse transformation, and you have been interviewed as a prospective consultant to supply the formula for the inversion. The nature of the processing of $F_c(\nu)$ and its purpose constitute proprietary information that has not been divulged to you. Take this assignment seriously; the financial compensation is very much higher than you are used to. Compose a competent professional report. ▷

44. A transparent fiber made of silica is lightly doped with germanium dioxide so that the refractive index n has a value n_0 on the longitudinal z-axis but drops off with radius r according to $n = n_0(1 - r^2/h^2)$, where h is a distance large compared with the radius of the fiber. The value of n_0 is only one or two percent larger than the value (around 1.55) for silica. A light ray having entered the fiber at $r = 0, z = 0$, makes an angle i with the z-axis. Show that the ray follows the sinusoidal curve

$$r = \frac{2L \tan i}{\pi} \sin \frac{\pi z}{2L}$$

where $L = \pi h / 2\sqrt{2n_0}$ is the distance from $z = 0$ to the first turning point of the ray. ▷

45. **Inverse function.** Show that the function $1/\pi x$ has $-1/\pi x$ as its convolutional inverse in the sense that $(1/\pi x) * (-1/\pi x) = \delta(x)$.

14

The Laplace Transform

Hitherto we have taken the variables x and s to be real variables. Now, however, let t be a real variable and p a complex one, and consider the integral

$$\int_{-\infty}^{\infty} f(t)e^{-pt}\,dt.$$

This is known as the (two-sided) Laplace transform of $f(t)$ and will be seen to differ from the Fourier transform merely in notation. When the real part of p is zero, the identity with the Fourier transform (as interpreted with noncomplex variables) is complete. In spite of this, however, there is a profound difference in application between the two transforms.

Alternative definitions of the Laplace transform include

$$\int_{0+}^{\infty} f(t)e^{-pt}\,dt,$$

which may be referred to as the one-sided Laplace transform, and

$$p\int f(t)e^{-pt}\,dt,$$

which may be referred to as the p-multiplied form.

The one-sided transform arises in the analysis of transients, where $f(t)$ comes into existence following the throwing of a switch at $t = 0$. However, if we deem that $f(t) = 0$ for $t < 0$, such cases are seen to be included in the two-sided definition. It is not always stressed that the lower limit of the integral defining the one-sided Laplace transform is $0+$; indeed in practice it is normally written as 0. One must remember that

$$\int_{0}^{\infty} f(t)e^{-pt}\,dt \qquad \text{usually means} \qquad \lim_{h\to 0}\int_{|h|}^{\infty} f(t)e^{-pt}\,dt.$$

Many writers have adopted the p-multiplied form for the sake of continuity with the operational notation of Heaviside, whose early success with the differential equations of electrical transients stimulated long-continuing development in transform theory. The admittance operator $Y(p)$, where p is interpreted as the time-derivative operator, turns out to be the p-multiplied Laplace transform of the response to a unit voltage step.

The p-multiplied transform is now dying out, and so the connection with a good deal of electrotechnical and applied-mathematical literature has been disturbed. On the other hand, by omitting the p, one can have exact correspondence with the Fourier transform. Furthermore, one can preserve the compactness of routine operational solutions (which save a line or two), by dealing in terms of the *impulse response* instead of, as formerly, the step response. Thus in the operational relation

$$i(t) = Y(p)v(t)$$

between an applied voltage waveform $v(t)$ and a response current $i(t)$, where p is interpreted as the time-derivative operator, $Y(p)$ proves to be the not-p-multiplied two-sided Laplace transform of the response to a unit voltage impulse. This operational formula is often also true for the one-sided transform.

The inversion formula for regaining $f(t)$ from its Laplace transform has the sort of symmetry possessed by the inverse Fourier transformation. However, the variable p, unlike ω, is normally complex, so that a contour of integration in the complex plane of p must be specified. Thus

$$f(t) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} F_L(p)e^{pt}\,dp,$$

where
$$F_L(p) = \int_{-\infty}^{\infty} f(t)e^{-pt}\,dt,$$

and c is a suitably chosen positive constant.

Advantages of the Laplace transform over the Fourier transform for handling electrical transient and other problems are often quoted in books. The essential advantage of the Fourier transform is its physical interpretability—as a spectrum, a diffraction pattern, and so on. Laplace transforms are not so interpretable, and once we take Laplace transforms of an equation we retain only a mathematical, not a physical, grasp of its meaning.

In general, however, the Laplace transform is used for initial-value problems, such as transients in dynamical systems, and is not used at all in a broad class of problems included in the present work.

For basic texts see Doetsch (1943) and van der Pol and Bremmer (1955), and for extensive lists of Laplace transforms see Erdélyi et al. (1954) and Roberts and Kaufman (1966). McCollum and Brown (1965) have provided an exhaustive compilation of transforms arising in electric circuit applications.

CONVERGENCE OF THE LAPLACE INTEGRAL

The definition integral tells us that we are to multiply a given function $f(t)$ by e^{-pt}, where p is any complex number, and integrate from $-\infty$ to $+\infty$. If we consider for a moment only values of p which are real, we can graph the product $f(t)e^{-pt}$ for a sequence of values of p (see Fig. 14.1). When $p = 0$, the product is equal to $f(t)$ itself; hence the Laplace transform at $p = 0$ is equal to the infinite integral of $f(t)$; that is,

$$F_L(0) = \int_{-\infty}^{\infty} f(t)\, dt.$$

The areas under the successive curves labeled $p = -0.1, -0.2$ give $F_L(-0.1)$, $F_L(-0.2)$. Clearly, as p goes more and more negative, e^{-pt} represents an ever-larger factor by which the values of $f(t)$ for positive values of t have to be multiplied. There may be a limit, as suggested by the figure, which p must exceed if the integral is to remain finite. Let this limiting value of p be α. When p has an imaginary part, it can be shown that the integral will not be finite if the real part of p falls below this same real number α. Similarly, if positive real values of p are considered, it is possible that the area under the product may become infinite to the left of the origin. Of course, if $f(t) = 0$ for $t < 0$ this cannot happen, nor can it happen if the definition integral is one-sided. But in general there will be a real number β which p may not exceed, and as before it can be shown in the complex case that the real part of p is what counts.

Laplace transforms therefore may not exist for all values of p, and in the typical case there is a range specified by

$$\alpha < \operatorname{Re} p < \beta$$

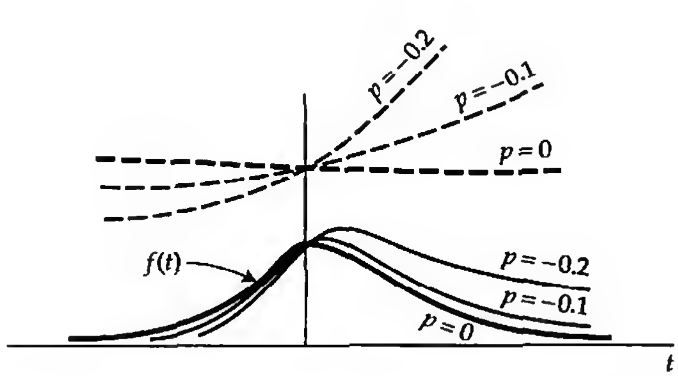


Fig. 14.1 The broken curves show e^{-pt}, and the full curves $e^{-pt}f(t)$, for certain real values of p.

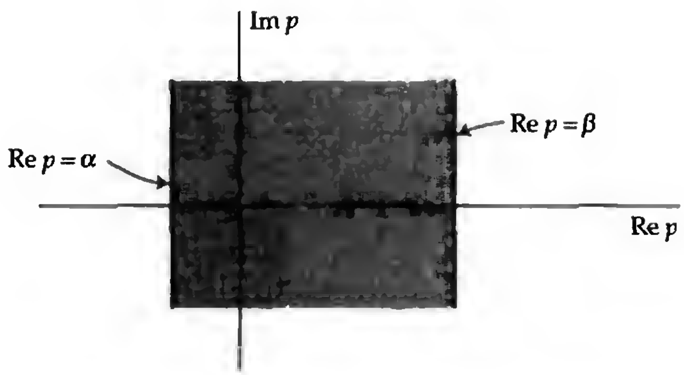


Fig. 14.2 **Complex plane of p showing the strip of convergence.**

in which the definition integral converges. This means that in the complex plane of p, convergence occurs in a strip (see Fig. 14.2). It is not necessary that convergence occur on the imaginary axis of p, as happens in the figure, or that it should occur at all.

Consider, for example, $f(t) = e^t$. This elementary function, widely used in circuit analysis and other subjects studied by means of the Laplace transform, does not possess a convergent Laplace integral if Re $p > 1$, or if Re $p < 1$. In this case the strip of convergence has contracted to a line: the integral converges only where Re $p = 1$, and even then not exactly at $p = 1$.

The function $\exp[tH(t)]$ would require Re $p < 0$ for the left-hand part of the integral to converge; but convergence of the right-hand side requires Re $p > 0$; hence there can be convergence for no value of p. On the other hand, $\exp[-tH(t)]$ can converge on the right-hand side if Re $p > -1$, hence there is a strip of convergence such that $-1 < \text{Re}\, p < 0$.

To summarize the convergence situation let us note that the left-hand limit α is determined by the behavior of the right-hand part of $f(t)$. The more rapidly $f(t)$ dies away with increasing t, the further to the left α may be, and if $f(t)$ falls to zero and remains zero, or diminishes as $\exp(-t^2)$ then α recedes to $-\infty$.

Similarly, the right-hand limit β depends on the left-hand half of $f(t)$, receding to $+\infty$ if, for example, $f(t)$ is zero to the left of some point. Examples illustrating the various possibilities are given in Fig. 14.3.

THEOREMS FOR THE LAPLACE TRANSFORM

Most of the theorems given for the Fourier transform are restated in Table 14.1 for the Laplace transform. To a large extent the theorems are similar, the main additional item concerning the strip of convergence.

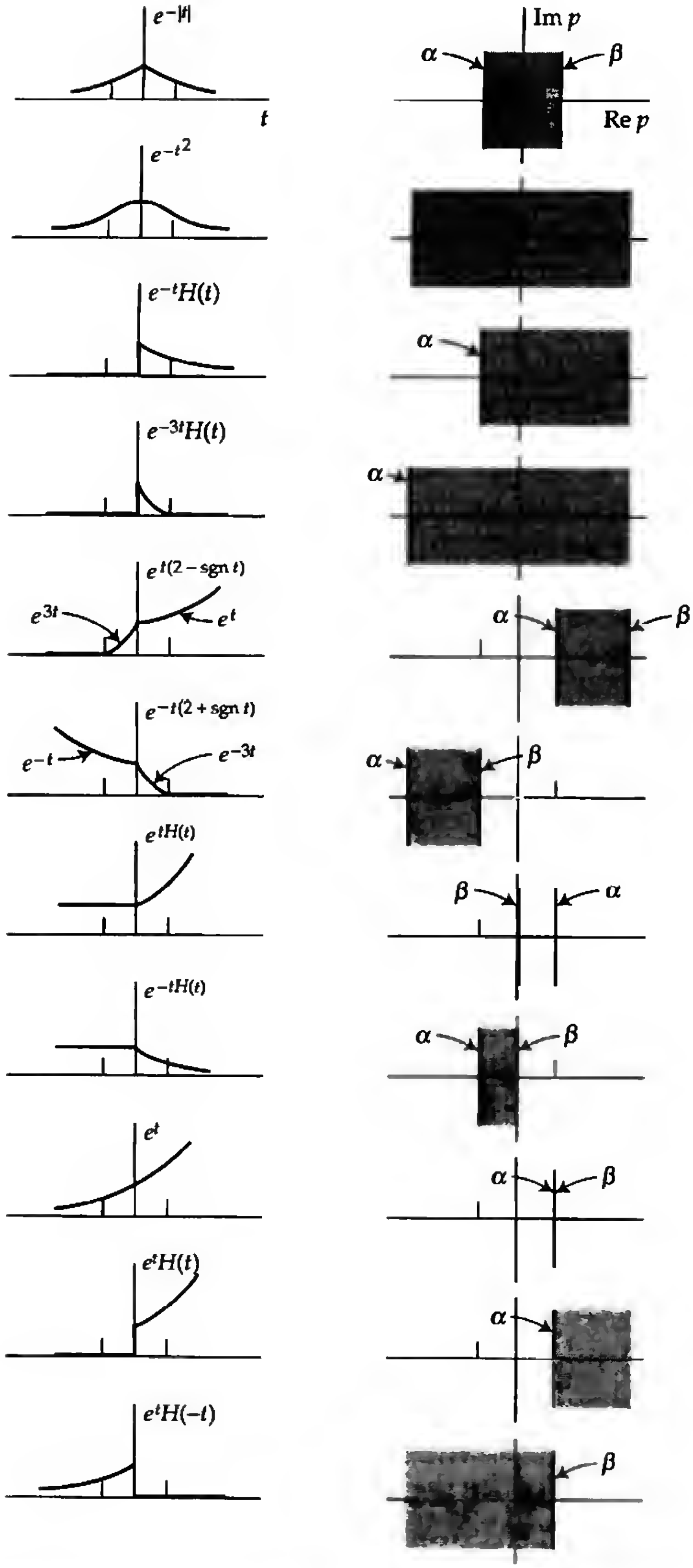


Fig. 14.3 **Some simple functions and their strips of convergence.**

TABLE 14.1
Theorems for the Laplace transform

Theorem	$f(t)$	$F(p)$	Strip of convergence
Similarity	$f(at)$	$\frac{1}{\lvert a\rvert}F\left(\frac{p}{a}\right)$	$\alpha < a^{-1}\,\mathrm{Re}\,p < \beta$
Addition	$f_1(t) + f_2(t)$	$F_1(p) + F_2(p)$	$\max(\alpha_1,\alpha_2) < \mathrm{Re}\,p < \min(\beta_1,\beta_2)$
Shift	$f(t-\tau)$	$e^{-\tau p}F(p)$	$\alpha < \mathrm{Re}\,p < \beta$
Modulation	$f(t)\cos\omega t$ $e^{-at}f(t)$	$\frac{1}{2}F(p-i\omega) + \frac{1}{2}F(p+i\omega)$ $F(p+a)$	$\alpha < \mathrm{Re}\,p < \beta$ $\alpha - \mathrm{Re}\,a < \mathrm{Re}\,p < \beta - \mathrm{Re}\,a$
Convolution	$f_1(t) * f_2(t)$	$F_1(p)F_2(p)$	$\max(\alpha_1,\alpha_2) < \mathrm{Re}\,p < \min(\beta_1,\beta_2)$
Product	$f_1(t)f_2(t)$	$\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} F_1(s)F_2(p-s)\,ds$	$\alpha_1 + \alpha_2 < \mathrm{Re}\,p < \beta_1 + \beta_2,\ \alpha_1 < c < \beta_1$
Autocorrelation	$f(t) * f(-t)$	$F(p)F(-p)$	$\lvert\mathrm{Re}\,p\rvert < \min(\lvert\alpha\rvert,\lvert\beta\rvert)$
Differentiation	$f'(t)$	$pF(p)$	$\alpha < \mathrm{Re}\,p < \beta$ (often)
Finite difference	$f(t+\frac{1}{2}\tau)$ $-f(t-\frac{1}{2}\tau)$	$2\sinh\frac{1}{2}\tau p\,F(p)$	$\alpha < \mathrm{Re}\,p < \beta$
Integration	$\int_{-\infty}^{t} f(u)\,du$	$p^{-1}F(p)$	$\max(\alpha,0) < \mathrm{Re}\,p < \beta$
	$\int_{t}^{\infty} f(u)\,du$	$p^{-1}F(p)$	$\alpha < \mathrm{Re}\,p < \min(\beta,0)$
Reversal	$f(-t)$	$F(-p)$	$-\beta < \mathrm{Re}\,p < -\alpha$

TRANSIENT-RESPONSE PROBLEMS

Let a time-varying stimulus $V_1(t)$ elicit a response $V_2(t)$ from a linear time-invariant system. For example, $V_1(t)$ might be the voltage applied to the input of a linear time-invariant electrical network, and $V_2(t)$ might be the output voltage caused by $V_1(t)$. Then the relation between the two will often be conveniently expressible in the form of a linear differential equation with constant coefficients:

$$a_0V_1(t) + a_1\frac{d}{dt}V_1(t) + a_2\frac{d^2}{dt^2}V_1(t) + \dots$$
$$= b_0V_2(t) + b_1\frac{d}{dt}V_2(t) + b_2\frac{d^2}{dt^2}V_2(t) + \dots.$$

Such a differential equation is suitable for solution by the Laplace transformation. Multiply each term by $\exp(-pt)$ and integrate each term from $-\infty$ to ∞. Then

$$a_0\int_{-\infty}^{\infty}V_1(t)e^{-pt}\,dt + a_1\int_{-\infty}^{\infty}V'(t)e^{-pt}\,dt + \dots$$
$$= b_0\int_{-\infty}^{\infty}V_2(t)e^{-pt}\,dt + b_1\int_{-\infty}^{\infty}V_2'(t)e^{-pt}\,dt + \dots.$$

Using bar notation for Laplace transforms, that is,

$$\overline{V}_1(p) = \int_{-\infty}^{\infty} V_1(t)e^{-pt}\,dt,$$

and recalling that the Laplace transform of the derivative $V_1'(t)$ is $p\overline{V}_1(p)$, we have

$$a_0\overline{V}_1(p) + a_1p\overline{V}_1(p) + a_2p^2\overline{V}_1(p) + \ldots = b_0\overline{V}_2(p) + b_1p\overline{V}_2(p) + b_2p^2\overline{V}_2(p) + \ldots.$$

This equation relating the Laplace transforms of $V_1(t)$ and $V_2(t)$ is known as the *subsidiary equation*. It is not a differential equation; it is merely an algebraic equation and may readily be solved for $\overline{V}_2(p)$. Thus, where the denominator is not zero,

$$\overline{V}_2(p) = \frac{a_0 + a_1p + a_2p^2 + \ldots}{b_0 + b_1p + b_2p^2 + \ldots}\overline{V}_1(p).$$

This simple step typifies the essence of transform methods. By translating a problem to the transform domain one gets a simpler problem.

It will be found that partial differential equations in which there are two independent variables, such as x and t on a transmission line, are reduced to ordinary differential equations. Simultaneous differential equations are reduced to simultaneous algebraic equations, and so on.

In the present case, by the act of solving the trivial algebraic problem, we have solved the differential equation. It only remains to retranslate the result into the original domain.

If $V_1(t)$ is a simple conventional stimulus, and we shall refer later to cases in which it is not, then its Laplace transform $\overline{V}_1(p)$ may be recalled from the memory or from a small stock of transform pairs kept for reference. The reference list is eked out by theorems and one or two established procedures, for example, the expansion of ratios of polynomials in partial fractions. After calculating $\overline{V}_2(p)$ from $\overline{V}_1(p)$ by multiplication with the ratio of polynomials as indicated, one retransforms, by the same procedures, to find $V_2(t)$.

LAPLACE TRANSFORM PAIRS

It is rarely necessary to evaluate a Laplace transform by integration. Most frequently, the method is used on systems with a finite number of degrees of freedom, for which a quite small stock of transform pairs suffices. The key pair is

$$e^{-\alpha t}H(t) \supset \frac{1}{p+\alpha} \qquad -\mathrm{Re}\,\alpha < \mathrm{Re}\,p.$$

Taking α to be zero, we have

$$H(t) \supset \frac{1}{p} \qquad 0 < \mathrm{Re}\,p.$$

Taking $\alpha = i\omega$, we have

$$e^{-i\omega t}H(t) \supset \frac{1}{p + i\omega} \qquad 0 < \operatorname{Re} p$$

and

$$e^{i\omega t}H(t) \supset \frac{1}{p - i\omega} \qquad 0 < \operatorname{Re} p.$$

Adding the last two pairs, we have

$$\begin{aligned}\cos \omega t\, H(t) &\supset \frac{1}{2}\frac{1}{p - i\omega} + \frac{1}{2}\frac{1}{p + i\omega} \\ &= \frac{p}{p^2 + \omega^2} \qquad 0 < \operatorname{Re} p,\end{aligned}$$

and subtracting, we have

$$\begin{aligned}\sin \omega t\, H(t) &\supset \frac{1}{2i}\frac{1}{p - i\omega} - \frac{1}{2i}\frac{1}{p + i\omega} \\ &= \frac{\omega}{p^2 + \omega^2} \qquad 0 < \operatorname{Re} p.\end{aligned}$$

By applying the integration theorem to $H(t) \supset p^{-1}$ we find

$$tH(t) \supset \frac{1}{p^2} \qquad 0 < \operatorname{Re} p,$$

and the derivative theorem gives

$$\delta(t) \supset 1 \qquad -\infty < \operatorname{Re} p < \infty.$$

By the addition theorem, we have

$$(1 - e^{-\alpha t})H(t) \supset \frac{\alpha}{p(p + \alpha)} \qquad \max\,(0, -\operatorname{Re}\alpha) < \operatorname{Re} p.$$

By differentiating $\exp(-\alpha t)H(t) \supset (p + \alpha)^{-1}$ with respect to α, we have

$$te^{-\alpha t}H(t) \supset \frac{1}{(p + \alpha)^2} \qquad -\operatorname{Re}\alpha < \operatorname{Re} p.$$

Splitting α into real and imaginary parts, $\alpha = \sigma + i\omega$, we have

$$e^{-(\sigma \pm i\omega)t}H(t) \supset \frac{1}{p + \sigma \pm i\omega} \qquad -\sigma < \operatorname{Re} p,$$

whence

$$e^{-\sigma t} \cos \omega t\, H(t) \supset \frac{p + \sigma}{(p + \sigma)^2 + \omega^2} \qquad -\sigma < \operatorname{Re} p,$$

and

$$e^{-\sigma t} \sin \omega t\, H(t) \supset \frac{\omega}{(p + \sigma)^2 + \omega^2} \qquad -\sigma < \operatorname{Re} p.$$

TABLE 14.2
Some Laplace transforms

$f(t)$	$\int_{-\infty}^{\infty} f(t)e^{-pt}\,dt$	Strip of convergence
$e^{-\alpha t}H(t)$ $-e^{-\alpha t}H(-t)$	$\dfrac{1}{p+\alpha}$	$-\mathrm{Re}\,\alpha < \mathrm{Re}\,p$ $\mathrm{Re}\,p < -\mathrm{Re}\,\alpha$
$H(t)$	$\dfrac{1}{p}$	$0 < \mathrm{Re}\,p$
$tH(t)$	$\dfrac{1}{p^2}$	$0 < \mathrm{Re}\,p$
$\delta(t)$	1	all p
$\delta'(t)$	p	all p
$(1 - e^{-\alpha t})H(t)$	$\dfrac{\alpha}{p(p+\alpha)}$	$\max(0, -\mathrm{Re}\,\alpha) < \mathrm{Re}\,p$
$\cos \omega t\, H(t)$	$\dfrac{p}{p^2+\omega^2}$	$0 < \mathrm{Re}\,p$
$\sin \omega t\, H(t)$	$\dfrac{\omega}{p^2+\omega^2}$	$0 < \mathrm{Re}\,p$
$te^{-\alpha t}H(t)$	$\dfrac{1}{(p+\alpha)^2}$	$-\mathrm{Re}\,\alpha < \mathrm{Re}\,p$
$e^{-\sigma t}\cos \omega t\, H(t)$	$\dfrac{p+\sigma}{(p+\sigma)^2+\omega^2}$	$-\sigma < \mathrm{Re}\,p$
$e^{-\sigma t}\sin \omega t\, H(t)$	$\dfrac{\omega}{(p+\sigma)^2+\omega^2}$	$-\sigma < \mathrm{Re}\,p$
$e^{-\alpha\|t\|}$	$\dfrac{2\alpha}{\alpha^2-p^2}$	$-\mathrm{Re}\,\alpha < \mathrm{Re}\,p < \mathrm{Re}\,\alpha$
$\Pi(t)$	$\dfrac{\sinh \frac{1}{2}p}{\frac{1}{2}p}$	all p
$\Pi(t-\frac{1}{2})$	$\dfrac{1-e^{-p}}{p}$	all p
$\Lambda(t)$	$\left(\dfrac{\sinh \frac{1}{2}p}{\frac{1}{2}p}\right)^2$	all p
$\mathrm{III}(t)H(t+\frac{1}{2})$	$\dfrac{1}{1-e^{-p}}$	$0 < \mathrm{Re}\,p$
$\mathrm{I}_{\mathrm{I}}(t)$	$\sinh \frac{1}{2}p$	all p
$\mathrm{II}(t)$	$\cosh \frac{1}{2}p$	all p
$-H(-t)$	$\dfrac{1}{p}$	$\mathrm{Re}\,p < 0$
$\mathrm{sgn}\, t$	$\dfrac{2}{p}$	$\mathrm{Re}\,p = 0$
$\mathrm{III}(t)H(t)$	$\coth \frac{1}{2}p$	$0 < \mathrm{Re}\,p$

From the representative variety of familiar transient waveforms that are so readily derivable from

$$e^{-\alpha t}H(t) \supset \frac{1}{p+\alpha},$$

it is clear that the exponential waveform plays a fundamental role. As shown elsewhere, exponential response to exponential stimulus is synonymous with linearity plus time invariance. All the time functions listed are thus inherent in linear time-invariant systems and may be elicited by simple stimuli.

For reference, the examples are tabulated in Table 14.2, together with a few other examples that are characteristic of simple stimuli. These further examples may all be readily obtained by integration, or converted directly from the known Fourier transforms, by replacing $i2\pi f$ by p. The strip of convergence has to be considered separately.

NATURAL BEHAVIOR

In the transient-response problem described above, the answer $V_2(t)$ is the response elicited by the stimulus $V_1(t)$. The system may also, however, be capable of natural behavior that requires no stimulus for its maintenance. Usually, some prior stimulus has been applied to the network to excite the natural behavior (but not necessarily; the system may have been assembled from elements containing stored energy).

Evidently any natural behavior associated with the circulation of energy not injected by the stimulus $V_1(t)$ has to be added to $V_2(t)$ to give the total behavior.

The well-known procedure for finding the natural behavior amounts to solving for a response $V_2(t)$ when the stimulus $V_1(t) = 0$. Thus putting $\overline{V}_1(p) = 0$ in the subsidiary equation, we have

$$b_0 + b_1 p + b_2 p^2 + \dots = 0.$$

This is known as the *characteristic equation*, and it furnishes the set of complex roots $\lambda_1, \lambda_2, \dots,$ that define the natural modes of the system.

A variation $V_2(t) = \exp(\lambda_k t)$ will be seen on substitution in the differential equation to be a solution when $V_1(t) = 0$. A simultaneous combination of all the modes, with arbitrary strengths, expresses the natural behavior in its most general form: $\Sigma_k a_k \exp \lambda_k t$, provided the roots $\lambda_1, \lambda_2, \dots$ are all different.

If there is a repeated root, say for example that λ_3 equals λ_4, then nonexponential behavior $t \exp \lambda_3 t$ can occur. In a sense this is conceivable as a combination of exponential modes as follows. Consider a system in which λ_3 and λ_4 are almost equal, and then consider a sequence of systems in which λ_3 and λ_4 approach equality as attention progresses from one system to the next. The expression

$$a_3 e^{\lambda_3 t} + a_4 e^{\lambda_4 t}$$

denotes possible forms of behavior, where a_3 and a_4 are arbitrary strengths. Let a_3 be equal to a_4 in magnitude, but opposite in sign. Let $\lambda_4 = \lambda_3 + \delta\lambda$. Then the expression above becomes

$$a_3[e^{\lambda_3 t} - e^{(\lambda_3+\delta\lambda)t}],$$

which describes a humped waveform, rising from zero and than falling back; but as $\delta\lambda$ approaches zero the expression also goes to zero. Since, however, the coefficients a_k are arbitrary, and the modes may be evoked in any strength, we are quite entitled to amplify the faint behavior as $\delta\lambda$ diminishes. Suppose that as we move from system to system we make the coefficients a_3 and a_4 increase in inverse proportion to $\delta\lambda$. Then a sequence of possible behaviors is

$$\frac{e^{(\lambda_3+\delta\lambda)t} - e^{\lambda_3 t}}{\delta\lambda},$$

an expression which in the limit is the derivative of exp $\lambda_3 t$ with respect to λ_3. Therefore, in the limiting system, where λ_4 is equal to λ_3,

$$te^{\lambda_3 t}$$

is a possible time variation and any multiple thereof. It might seem that the merging modes would have to be hit extremely hard in order to produce this result, but in fact only a finite amount of energy is required. Hence, if repeated roots occur, derivatives of exp $\lambda_k t$ with respect to λ_k must be included in the expression for the most general natural behavior.

■

IMPULSE RESPONSE AND TRANSFER FUNCTION

Let $V_2(t) = I(t)$ when $V_1(t) = \delta(t)$, irrespective of the choice of the origin of t. Then $I(t)$ may be called the impulse response, and by the superposition property,

$$V_2(t) = \int_{-\infty}^{\infty} I(u)V_1(t-u)\,du.$$

Now if $V_1(t)$ is an exponential function,

$$V_1(t) = e^{pt},$$

where p is some complex constant, it follows that

$$\begin{aligned} V_2(t) &= \int_{-\infty}^{\infty} I(u)e^{p(t-u)}\,du \\ &= \left[\int_{-\infty}^{\infty} I(u)e^{-pu}\,du\right]e^{pt}. \end{aligned}$$

The result shows that the response to exponential stimulus is also exponential, with the same constant p, but with an amplitude given by the quantity in brackets. We call this quantity the *transfer function*, and it is seen to be the Laplace trans-

form of the impulse response. From the subsidiary equation given earlier we see that the transfer function is

$$\frac{a_0 + a_1 p + a_2 p^2 + \dots}{b_0 + b_1 p + b_2 p^2 + \dots}.$$

For lumped electrical networks, one can write the transfer function readily from experience with calculating a-c impedance. Hence, through the Laplace transformation, the impulse response also becomes readily accessible. Thus in the circuit of Fig. 14.4, the ratio of $V_2(t)$ to $V_1(t)$ when $V_1(t)$ varies as exp pt is, by inspection,

$$R\frac{\dfrac{\dfrac{1}{R+L_2p}}{\dfrac{1}{R+L_2p}+Cp}}{L_1p+\dfrac{(R+L_2p)\dfrac{1}{Cp}}{R+L_2p+\dfrac{1}{Cp}}} = R\frac{1}{R+(L_1+L_2)p+L_1RCp^2+L_1L_2Cp^3}.$$

By the method of partial fractions, this ratio of polynomials can be expressed in the form

$$\frac{A}{p+a}+\frac{B}{p+b}+\frac{C}{p+c}$$

and thus the impulse response is

$$(Ae^{-at}+Be^{-bt}+Ce^{-ct})H(t).$$

In problems where $V_2(t)$ must be found when $V_1(t)$ is given, two distinct courses must always be borne in mind: one is to proceed by multiplication with the transfer function in the transform domain; the other method is to proceed, by use of the impulse response, directly through the superposition integral.

A strange difference between the two procedures may be noted. The superposition integral

$$V_2(t_1) = \int_{-\infty}^{\infty} I(u)V_1(t_1-u)\,du$$

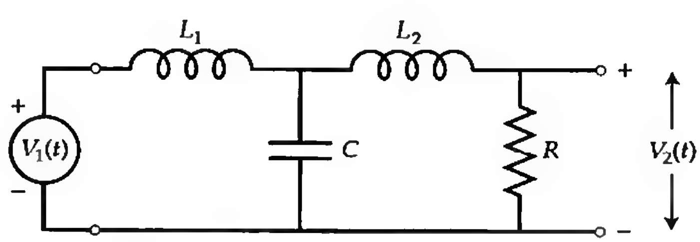


Fig. 14.4 **Find $V_2(t)$ when $V_1(t)$ is given.**

does not make use of values of $V_1(t)$ for $t > t_1$ during the course of calculating $V_2(t_1)$. (This is clear because of the fact that the integrand vanishes for negative values of u.) In other words, to calculate what is happening now, we need only know the history of excitation up to the present moment.

But the procedure in the transform domain normally makes use of information about the future excitation. Indeed, to calculate the Laplace transform of the excitation $V_1(t)$ we have to know its behavior indefinitely into the future.

As a rule, one obtains the same result for the response at $t = t_1$ irrespective of what $V_1(t)$ does after t_1 passes, but occasional trouble can be expected. For example, it is perfectly reasonable to ask what is the response of a system to $V_1(t) = \exp t^2[H(t)]$, and the superposition integral gives it readily, but the Laplace integral diverges because of troubles connected with the infinitely remote future. One way out of this dilemma is to assume that the excitation is perpetually on the point of ceasing; in other words, assume an excitation $\exp t^2[H(t) - H(t - t_1)]$.

INITIAL-VALUE PROBLEMS

Various types of initial-value problems may be distinguished. One type represents the classical problem of a differential equation plus boundary conditions. Thus the driving function is given for all[1] $t > 0$, but no information is provided about the prior excitation of the system. Consequently, some continuing response to prior excitation may still be present, and therefore extra facts must be furnished. The number of extra facts is the same as, or closely connected with, the number of modes of natural behavior, for the continuing response may be regarded as a mixture of natural modes excited by a stimulus that ceased at $t = 0$. Thus sufficient facts must be supplied to allow the strength of each natural mode to be found. These data often consist of the "initial" response and a sufficient number of "initial" derivatives. The word "initial" refers to $t = 0+$; the value at $t = 0$, which may seem to have more claim to the title, is frequently different. To deal with this problem, calculate the response due to a stimulus $V_1(t)$ that is equal to zero for $-\infty < t < 0$,[2] and equal to the given driving function for $t > 0$.[3] Of the many possible particular solutions of the differential equation, this one corresponds to the condition of no stored energy prior to $t = 0$. At $t = 0+$, this response function and its derivatives possess values that may be compared with the initial data. If they are the same, then there was no stored energy prior to $t = 0$, and the solution is complete. But if there are discrepancies, the differences must

[1] The exciting function $V_1(t)$ is given sometimes for $t \geq 0$ and sometimes for $t > 0$, but the statement of the problem, as in the example that follows, often does not say which.

[2] In this statement $-\infty$ means a past epoch subsequent to the assembly of a physical system which has since remained in the same linear, time-invariant condition.

[3] At $t = 0$, take a value equal to $\frac{1}{2}V(0+)$; or take any other *finite* value, since, as previously seen, there is no response to a null function.

be used to fix the strengths a_k of the natural modes by solving a small number of simultaneous equations of the form

$$V_2(0+) = \sum_k a_k e^{\lambda_k t}$$

$$V_2'(0+) = \sum_k a_k \lambda_k e^{\lambda_k t},$$

$$\ldots .$$

A much better method, which injects the boundary conditions in advance, depends on assuming that the *response*, not the stimulus, is zero for $t < 0$. Let it be required to solve the differential equation

$$b_0\hat{V}_2(t) + b_1\hat{V}_2'(t) + b_2\hat{V}_2''(t) + \ldots = \hat{V}_1(t)$$

for $\hat{V}_2(t)$, given $\hat{V}_1(t)$ for $t > 0$, and the initial values $\hat{V}_2(0+)$, $\hat{V}_2'(0+)$, ... The circumflex accents indicate that the functions are undefined save for positive t. We replace the differential equation by a new one for $V_2(t)$ where

$$V_2(t) = \begin{cases} \hat{V}_2(t) & t > 0 \\ 0 & t < 0. \end{cases}$$

The function $V_2(t)$ is thus defined for all t, and agrees with $\hat{V}_2(t)$, where the latter is defined. Since

$$V_2'(t) = \hat{V}_2(0+)\delta(t) + \begin{cases} \hat{V}_2'(t) & t > 0 \\ 0 & t < 0 \end{cases}$$

and
$$V_2''(t) = \hat{V}_2'(0+)\delta(t) + \hat{V}_2(0+)\delta'(t) + \begin{cases} \hat{V}_2''(t) & t > 0 \\ 0 & t < 0, \end{cases}$$

the new differential equation is

$$b_0V_2(t) + b_2V_2'(t) + b_2V_2''(t) + \ldots = V_1(t),$$

where
$$V_1(t) = [b_1\hat{V}_2(0+)\delta(t)] + [b_2\hat{V}_2'(0+)\delta(t) + b_2\hat{V}_2(0+)\delta'(t)] + \ldots + \begin{cases} \hat{V}_1(t) & t > 0 \\ 0 & t < 0. \end{cases}$$

This differential equation differs from the original one by the inclusion in the driving function of impulses necessary to produce the initial discontinuities of $V_2(t)$ and its derivatives. Taking Laplace transforms throughout, we have the subsidiary equation

$$b_0\overline{V}_2(p) + b_1p\overline{V}_2(p) + b_2p^2\overline{V}_2(p) + \ldots = \overline{V}_1(p).$$

Thus
$$\overline{V}_2(p) = \frac{\overline{V}_1(p)}{b_0 + b_1p + b_2p^2 + \ldots}$$

$$= \frac{[b_1\hat{V}_2(0+)] + [b_2\hat{V}_2'(0+) + b_2\hat{V}_2(0+)p] + \ldots + \int_{0+}^{\infty} \hat{V}_1(t)e^{-pt}\,dt}{b_0 + b_1p + b_2p^2 + \ldots}.$$

Inversion of this transform to the time domain gives $V_2(t)$; and the part of it referring to $t > 0$ gives the desired solution $\hat{V}_2(t)$, with built-in satisfaction of the initial conditions.

Consider the following example: Given the differential equation

$$\frac{d^2y}{dt^2} + \frac{3\,dy}{dt} + 2y = 2,$$

solve for $y(t)$ ($t > 0$), subject to the initial conditions $y(0) = 1$, $y'(0) = 0$. (Although this is customary notation, $y(0)$ really stands for $y(0+)$, as we have noted.) It may be observed that the driving function is 2, independent of time. However, it is not explicitly stated whether the driving function is equal to 2 for $t > 0$, or for $t \geq 0$. Let

$$Y(t) = \begin{cases} y(t) & t > 0 \\ 0 & t < 0, \end{cases}$$

as shown in Fig. 14.5. Then

$$Y'(t) = y(0+)\,\delta(t) + \begin{cases} y'(t) & t > 0 \\ 0 & t < 0 \end{cases}$$

and
$$Y''(t) = y'(0+)\,\delta(t) + y(0+)\,\delta'(t) + \begin{cases} y''(t) & t > 0 \\ 0 & t < 0, \end{cases}$$

and the new differential equation for $Y(t)$ is

$$Y''(t) + 3Y'(t) + 2Y = \delta'(t) + 3\delta(t) + 2H(t).$$

The subsidiary equation is

$$p^2\overline{Y}(p) + 3p\overline{Y}(p) + 2\overline{Y}(p) = p + 3 + (2/p),$$

whence
$$\overline{Y}(p) = \frac{p + 3 + (2/p)}{p^2 + 3p + 2}$$
$$= \frac{1}{p}$$

and hence
$$Y(t) = H(t) \text{ for all } t.$$

Thus finally, for the limited region $t > 0$, the desired solution is

$$y(t) = 1.$$

It is now instructive to study this example from the standpoint of investigating the prior excitation, which was not divulged in the statement of the problem, and in lieu of which the initial conditions were given. For the sake of concreteness, let it be observed that

$$L\frac{d^2Q}{dt^2} + R\frac{dQ}{dt} + \frac{Q}{C} = V(t)$$

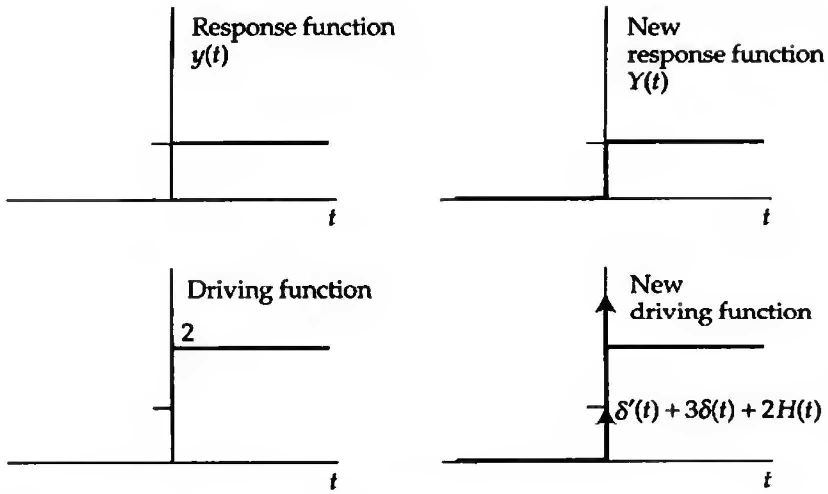


Fig. 14.5 An initial-value problem (left) and its equivalent (right).

is the differential equation for the charge Q on the capacitor of a series LRC circuit across which there is a voltage $V(t)$. If $L = 1$, $R = 3$, $C^{-1} = 2$, we deduce from the previous work that a Q of one coulomb, constant for $t > 0$, is compatible with an applied voltage of 2 volts ($t > 0$), and with the conditions $Q(0+) = 1$ and $Q'(0+) = 0$.

Unless the circuit was assembled from elements with energy stored in them, the capacitor charge must have been drawn from the voltage source. We can easily find what the voltage source must have done by making some simple supposition. Suppose, for example, that the charge was placed instantaneously at $t = 0$; that is, take

$$Q(t) = H(t).$$

Then, from the differential equation,

$$V(t) = \delta'(t) + 3\delta(t) + 2H(t).$$

The impulsive parts force a current impulse through the inductance and resistance, placing the charge on the capacitor, and the voltage step holds it there.

We have just solved a problem in which a complete history of excitation is given and no supplementary initial conditions enter. By now suppressing the information about the drastically impulsive behavior of $V(t)$ at $t = 0$ and introducing in its stead the fact that at $t = 0+$ the capacitor was seen to be already charged, with zero rate of change, we return to the original problem.

There are many other ways in which the capacitor could have been given its initial charge. In particular, in this problem, the charging could have been done in the same way but earlier, as described by

$$Q(t) = H(t + 1)$$
$$V(t) = \delta'(t + 1) + 3\delta(t + 1) + 2H(t + 1).$$

The positive-t part of these quantities would also satisfy the initial-condition problem, but $Q(0)$ and $V(0)$ would not be the same as before.

The peculiarity of the systematic method given for incorporating initial conditions is that impulsive terms are added to the driving function so as to inject all requisite initial energy instantaneously at $t = 0$.

SETTING OUT INITIAL-VALUE PROBLEMS

When the principle of the Laplace transform solution is understood, fewer steps in setting out are required. From the differential equation and initial conditions the subsidiary equation is written directly by inspection. Thus, given

$$\frac{d^2y}{dt^2} + \frac{3\,dy}{dt} + 2y = \varphi(t),$$

we write

$$p^2\,{}_0\bar{y}(p) + 3p\,{}_0\bar{y}(p) + 2\,{}_0\bar{y}(p) = [y'(0+) + y(0+)p] + [3y(0+)] + {}_0\bar{\varphi}(p),$$

where
$$_0\bar{y}(p) \subset \begin{cases} y(t) & t > 0 \\ 0 & t < 0, \end{cases}$$

and
$$_0\bar{\varphi}(p) \subset \begin{cases} \varphi(t) & t > 0 \\ 0 & t < 0. \end{cases}$$

Since these transforms refer to functions that are zero for $t < 0$, they are equal to the one-sided transforms of the functions $y(t)$, $\varphi(t)$. Early emphasis on this type of problem accounts for the prevalence of the one-sided transform.

SWITCHING PROBLEMS

Two types of problems have been described which are suited for handling by Laplace transforms:

1. Transient response of a linear, time-invariant system to a driving function given for all t (includes problem of natural behavior on taking driving function to be zero).
2. Initial-value problems where information on the driving function for $t > 0$ is supplemented by initial data regarding the response.

In a third type of problem the time invariance is violated in a simple way by switching. It is important to realize that throwing a switch changes an electric circuit to a different circuit, as regards application of the linear, time-invariant theory. For example, when switching places a battery voltage across a resistor, a new

circuit comes into existence. This simple switching problem is quite different in kind from the transient-response problem in which a voltage source in series with a resistor develops a voltage $H(t)$. It is true that in each case a step function of current flows in the resistor, but the second case involves a time-invariant circuit whose resistance is the same before $t = 0$ and after $t = 0$. There is an equivalence between the two problems, but only the second is amenable to the treatment already given for linear, time-invariant systems.

Of course, after the switch is thrown, both systems are identical. Consequently, if initial values of response are furnished ("initial" means after the switch has closed), together with details of the subsequent excitation, if any, the problem can be handled as an initial-value problem; for in those problems one does not inquire into $t < 0$.

More commonly, however, preinitial conditions (at $t = 0-$) are furnished, or the previous history of excitation is given. The technique for handling switching problems is to convert to an equivalent problem concerning a time-invariant circuit, as in the simple case of the battery and resistor.

A voltage $V_1(t)$, given for all t, is applied to a circuit; at $t = 0$ an open switch somewhere in the circuit closes. To find the response $V_2(t)$, at some place, we can proceed as before for times preceding the throwing of the switch. We could also easily calculate what the response to $V_1(t)$ would have been if the switch had always been closed. In fact we expect that this response will become apparent after the switch has been closed long enough, and the circuit has "forgotten" the effects of the preswitching excitation. To find the response during the transition following $t = 0$ is the present problem.

Across the terminals of the open switch a certain voltage exists, and it could be calculated as a transient response to the given driving function $V_1(t)$. Let it be $V_s(t)$. If a voltage source having zero internal impedance were connected across the switch, and if its voltage varied in exact agreement with $V_s(t)$, no current would flow in it. Since no current was flowing through the open switch, the voltage source could thus be connected without disturbing the behavior of the circuit. An important change has, however, been effected, for the zero-impedance connection now introduced across the switch has brought the circuit, for negative values of time, into identity with the circuit which was to be created at $t = 0$ by throwing the switch. The throwing of the switch can be fully simulated by the introduction of a voltage source $V_s(t)H(-t)$. This can be split into two terms $V_s(t) - V_s(t)H(t)$. The response associated with the first term is already known; it is the response, for all t, of the open-switch system to $V_1(t)$. (The superposition of the responses of the closed-switch system to $V_1(t)$ and $V_s(t)$ would give the same result.) The response associated with the second term, $-V_s(t)H(t)$, gives the desired transition term that has to be added to the response of the open-switch system to get the answer to the problem.

In problems where a switch is opened and any current through it thereby interrupted, one applies the principle of duality and introduces a suitable current generator.

BIBLIOGRAPHY

Doetsch, G.: "Theory and Application of the Laplace Transformation," Dover Publications, New York, 1943.

Erdélyi, A., ed.: "Tables of Integral Transforms," McGraw-Hill, New York, 1954.

McCollum, P. A., and B. F. Brown: "Laplace Transforms, Tables and Theorems," Holt, Rinehart and Wilson, New York, 1965. Provides a well-organized list of ratios of polynomials with denominators up to degree 5 and their transforms.

Roberts, G. E., and H. Kaufman: "Table of Laplace Transforms," W.B. Saunders, Philadelphia, 1966.

van der Pol, B., and H. Bremmer: "Operational Calculus Based on the Two-sided Laplace Integral," Cambridge University Press, Cambridge, UK, 1955.

PROBLEMS

1. State the strip of convergence for the following functions:

$$H(t) \qquad \Pi(t) \qquad \exp(-t^{\frac{1}{2}})H(t) \qquad (1+t^2)^{-1}$$
$$(1+t^3)^{-1}H(t) \qquad (\log t)H(t) \qquad \operatorname{sgn} t \qquad t^{-1}\Pi(t)$$
$$t^{-1} \qquad \delta(t) \qquad (1+e^t)^{-1} \qquad (1+t^3)^{-1}$$

2. Find the Laplace transforms of the following functions, stating the strip of convergence:

$$(e^{-t}-e^t)H(t) \qquad e^{-|t|} \qquad (e^t-e^{-t})H(-t)$$

3. Derive the Laplace transforms of $\Pi(t)$, $\Pi(t-\frac{1}{2})$, and $\Lambda(t)$, stating the strip of convergence. ▷

4. Find the current that flows in a series-LR circuit in response to a voltage $\exp(-|t|)$.

5. From one coil, one capacitor, and one resistor, 17 different two-terminal networks can be made. Tabulate the step and impulse responses of all these. ▷

6. Find the current in a series-LR circuit in response to a voltage $\exp(t) \sin t \, H(t)$.

7. A current impulse passes through a series-LR circuit. What is the voltage across the circuit?

8. The input impedance to a short-circuited length of cable is $75 \tan(10^{-8}f)$. A voltage generator of zero internal impedance applies an impulse. What current flows in response? ▷

9. A voltage $\sin \omega t \, H(t)$ is applied to a series-LR circuit. Show that the response current is

$$\frac{L\omega e^{-Rt/L} - L\omega \cos \omega t + R \sin \omega t}{R^2 + L^2\omega^2} H(t).$$

10. A voltage $\cos \omega t\, H(t)$ is applied to a series-RC circuit. Show that the response current is

$$\frac{R\omega^2C^2 \cos \omega t - \omega C \sin \omega t + R^{-1}e^{-t/RC}}{1 + \omega^2C^2R^2} H(t). \triangleright$$

11. A voltage $\exp(-at)H(t)$ is applied to a series-LR circuit. Show that the response current is

$$\frac{e^{-at} - e^{-Rt/L}}{R - aL} H(t).$$

12. A series-LR circuit is shunted by a capacitor C, and a ramp voltage $tH(t)$ is applied. Show that the response current is

$$\left[C + \frac{t}{R} - \frac{L}{R^2} + \frac{L}{R^2}e^{-Rt/L}\right] H(t).$$

13. Alternating voltage drives d.c.. An alternating voltage $\sin \omega t\, H(t)$ is applied to a circuit, and direct current is observed. What is the circuit? $\triangleright$

14. Battery delivers alternating current. A steady voltage is applied to a circuit, and a current $\sin \omega t\, H(t)$ results. What is the circuit? $\triangleright$

15. A capacitor C carries a charge Q. At $t = -1$ a switch is closed, allowing the capacitor to discharge through a resistor R. At $t = 0$ the switch is opened. Make a graph showing the current flow during the interval $-2 < t < 1$.

16. A resistor and an uncharged capacitor are connected in parallel to form a two-terminal network. At $t = -1$ a battery is switched across the network, and at $t = 0$ the switch is opened. Find the current in the resistor and capacitor for all t.

17. The voltage source applies a voltage $\exp(t)H(t)$ (see Fig. 14.6). At $t = 1$ the switch closes. Find the current through the voltage source during the interval $-\infty < t < \infty$, assuming the capacitor to have been initially uncharged.

18. At $t = -1$ capacitors C_1 and C_2 of Fig. 14.7 were seen to be discharging through the circuit containing the resistor R. At $t = 0$ the switch closed. Find the current through C_1 for all t.

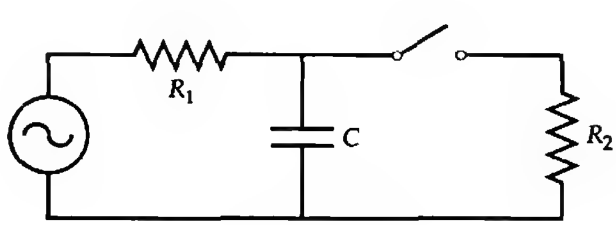


Fig. 14.6

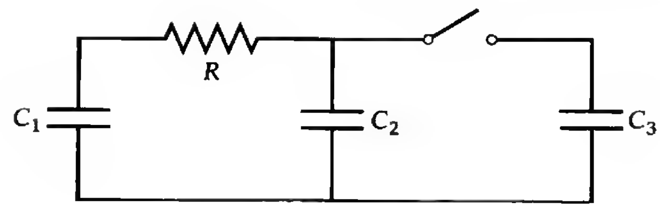


Fig. 14.7

19. Solve the following differential equations for $y(t)(t > 0)$, subject to initial values $y(0+)$, $y'(0+), \ldots$.
(a) $y'(t) + y(t) = 1$
Answer: $y = 1 + [y(0+) - 1]e^{-t}$
(b) $y''(t) + 5y'(t) + 6y(t) = 6$
Answer: $1 + [3y(0+) + y'(0+) - 3]e^{-2t} + [2 - y'(0+) - 2y(0+)]e^{-3t}$
(c) $y''(t) + 2y'(t) + y(t) = 2 \sin t$
Answer: $y = [1 + y(0+) + y'(0+)]te^{-t} + [1 + y(0+)]e^{-t} - \cos t$

20. For each of the above problems derive an electric circuit obeying the same equation. Deduce driving functions for all t that would produce the responses quoted for $t > 0$.

21. Strip of convergence. Find four different time functions all having $[(p - 2)(p - 1)(p + 1)]^{-1}$ as their Laplace transform. For each time function state the strip of convergence. ▷

22. Show that the Laplace transform of the staircase function $H(t) + H(t - T) + H(t - 2T) + \ldots$ is given by

$$\frac{1}{p(1 - e^{-Tp})} \qquad 0 < \operatorname{Re} p.$$

23. A voltage $v(t)H(t)$ is applied to a circuit that already contained energy before $t = 0$. As no information about the stimulus has been suppressed, no compensating postinitial data are needed. However, in order to get the total behavior $f(t)$ it will be necessary to know about the natural behavior that is going on independently. It may be that the full history of earlier energy injection is unavailable, but it will suffice to know the situation immediately prior to the application of the voltage at $t = 0$. For example, the energy stored in each inductor and capacitor at $t = 0-$ might be given; but suppose here that a sufficient number of preinitial values of the behavior are given: $f(0-), f'(0-), f''(0-), \ldots$. Show that the *total* behavior can be conveniently calculated by means of a special form of the Laplace transform defined by

$$F_-(p) = \int_{0-}^{\infty} f(t)e^{-pt}\,dt.$$

Work out the theorem for this transform, deducing, for example, that the derivative theorem is $f'(t) \supset pF_-(p) - f(0-)$. ▷

24. Convergence of Laplace integral. Find the region of convergence of the Laplace transform of $t^{-1}H(t-1)$.

25. Applying two-sided *LT* to initial value problem. Show how to solve $\dot{y} + y = 0$ subject to initial condition $y = y(0+)$ by the two-sided Laplace theorem. ▷

26. One-sided Laplace transform. A series LR circuit obeys the differential equation

$$LI'(t) + RI(t) = V(t).$$

Find out what current flows when the voltage is turned on at $t = 1$. Assume that no stored magnetic energy was previously present. ▷

27. Decaying echo train. The impulse response of a certain system is a regular train of impulses of exponentially decaying strength. Calculate the transfer function. ▷

28. Short-circuited transmission line. The input impedance to a two-terminal network is $iR \tan 2\pi Tf$. Show that the current that flows in response to an applied voltage impulse is

$$I(t) = R^{-1}\left[\delta(t) + 2\sum_{1}^{\infty} \delta(t - 2nT)\right]$$

and that the step response is

$$A(t) = R^{-1}\left[H(t) + 2\sum_{1}^{\infty} H(t - 2nT)\right].$$

29. Triggered clock-pulse generator. The network of the previous problem is placed in parallel with a resistance R. Show that the current that flows in response to a voltage impulse is given by

$$I(t) = 2R^{-1}\sum_{0}^{\infty} \delta(t - 2nT).$$

30. A suspicious transform. Authors listing Laplace transform pairs often indicate that $\text{III}(t)H(t)$ has Laplace transform $\coth\frac{1}{2}p$, $(0 < \text{Re}\, p)$. The corresponding Fourier transform is $\coth\frac{1}{2}i\omega$ or $-i\cos\pi f$, which is an odd function of f. Now $\text{III}(t)H(t)$ is not an odd function, having an even part as well as an odd part. Therefore the implied Fourier pair cannot be correct. (*a*) Show that when the missing even part of the Fourier transform is included, the correct Fourier transform pair is $\text{III}(t)H(t) \supset \frac{1}{2}\text{III}(f) - i\cot\pi f$. ▷

Photo A The Microwave Spectroheliograph.

The antenna array illustrated achieved a beamwidth of 3.1 arcminutes (Bracewell, R. N. and G. Swarup, "The Stanford Microwave Spectroheliograph Antenna, a Microsteradian Pencil Beam Interferometer," *IEEE Trans. Ant. Propag.*, vol. AP-9, pp. 22–30, 1961) and was used for making daily temperature maps of the Sun at a wavelength of 9.1 cm for 11 years. The east-west arm, comprising 16 steerable 3-m paraboloids on equatorial mounts spaced $d = 7.5$ m is representable as a convolution between the aperture distribution over a single paraboloid and the configuration function $\mathrm{III}(x/d)\Pi(x/16d)\delta(y)$, whose two-dimensional Fourier transform is $16d^2\mathrm{III}(dl) * \operatorname{sinc} 16dl$. Signals received from the Sun by the two arms are multiplied together; hence by the convolution theorem, the field reception pattern is proportional to $[\mathrm{III}(dl) * \operatorname{sinc} 16dl][\mathrm{III}(dm) * \operatorname{sinc} 16dm]$. By this clear Fourier-based reasoning we see that the overall pattern consists of a sharp beam of the form $\operatorname{sinc} 16dl \operatorname{sinc} 16dm$ pointed at the zenith, each such beam being replicated in a square pattern of spacing λ/d, or 42 arcminutes. From beam center to the first null is $\lambda/16d$ or 2.6 arcminutes.

In another Fourier approach to analysing the antenna pattern, each arm configuration may be broken into pairs of the form $\sum[\delta(x + \frac{1}{2}nd) + \delta(x - \frac{1}{2}nd)]$, $n = 1, 3, 5, 7$. Then the pattern is seen as the sum of four cosines that directly express the full pattern of one arm, including replication. In a subsequent development, the east-west arm was combined with a second interferometer extending a further 51 m to the west, to reduce the beamwidth to 50 arcseconds, as demonstrated on the radio star Cygnus A, thus taking the angular resolution of an antenna system beyond the visual acuity of the human eye.

For correct adjustment, the waveguide transmission lines from each antenna to the central junction need to be equal to ± 1 mm in an overall distance of 51 m, or 2 parts in 10^5, a precision characteristic of geodetic survey. To achieve this in an environment of varying temperature and humidity a fluorescent tube was placed in the feed horn of each antenna and switched on and off at 400 Hz to produce ionization capable of reflecting a faint modulated echo back to a central signal generator. Drifts in the phase path to any individual antenna could then be measured in the presence of strong unmodulated reflections from the branched waveguide system as a whole. This technique (G. Swarup and K. S. Yang, "Phase Adjustment of Large Antennas," *IEEE Trans. Ant. Propag.*, vol. AP-9, pp. 75–81, 1961) is the ancestor of various later related means for automatic phase-calibration and adjustment and is analogous to the development of telescopes with adaptive optics.

From the punched-tape output of the receiver a teletypewriter displayed a real-time sun map that was distributed by telephone. An analogue optical-electronic device provided the spherical trigometrical data for the day that controlled the varying declination and time of start required for successive rows of the sun map. The system was the first radio telescope to produce its output in camera-ready publishable form.

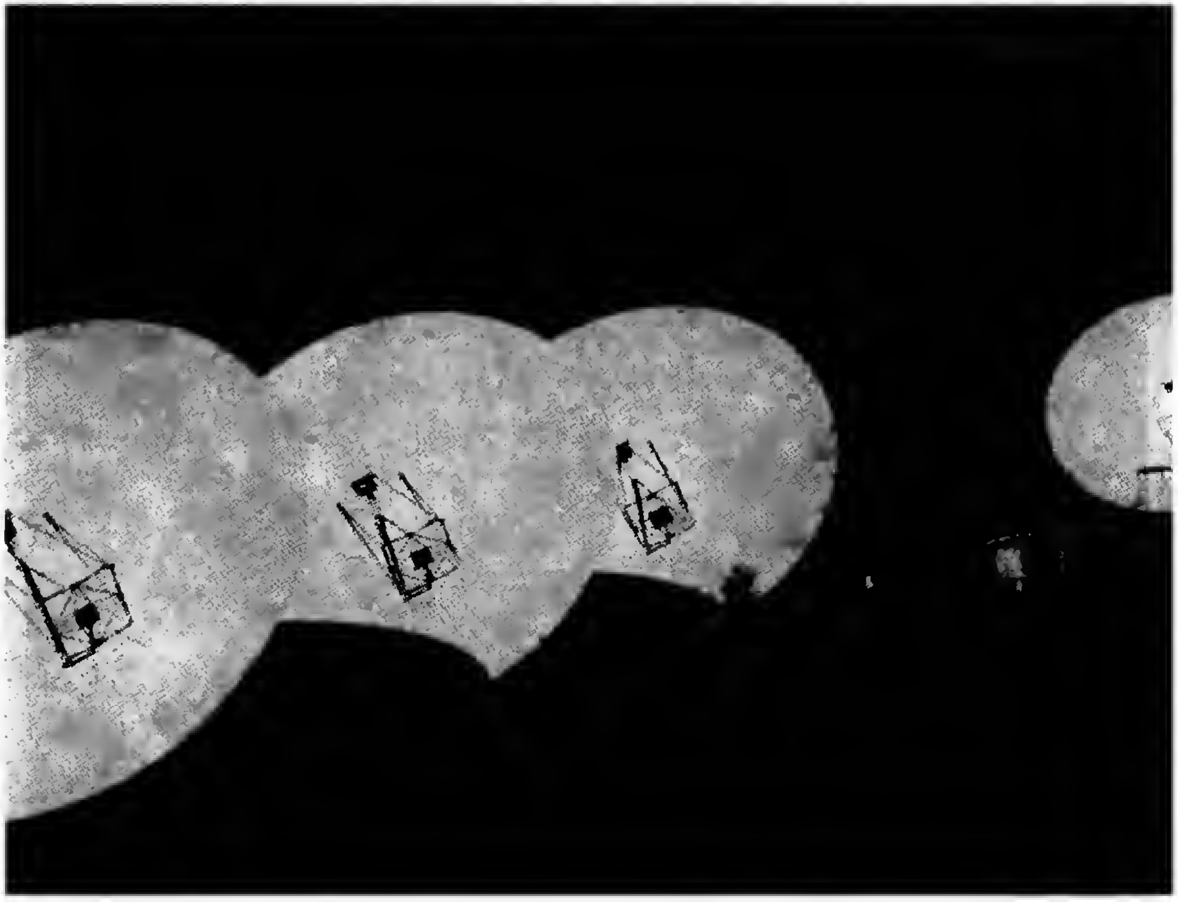


Photo B A Five-Element Radio Telescope.

Five equatorially mounted paraboloids, 18.3 m in diameter, spaced in units of $D = 22.9$ m, combine to produce fan beams with east-width beamwidth of 18.8 arcseconds at a wavelength of 2.8 cm (Bracewell, R. N., R. S. Colvin, L. R. D'Addario, C. J. Grebenkemper, K. M. Price, and A. R. Thompson, "The Stanford Five-Element Radio Telescope," *Proc. IEEE*, vol. 61, pp. 1249–1257, 1973). The fan beams are spaced 4.2 arcminutes. Earth rotation causes radio sources that are in the 7 arcminute beam to be scanned by the fan beams. The fan beams remain stationary on the sky in planes parallel to the meridian even as the antennas track the radio source, because the configuration function $\delta(x - D) + \delta(x - 2D) + \delta(x - 3D) + \delta(x - 7D) + \delta(x - 10D)$ is stationary. The scanning angle changes in time allowing the brightness distribution over a source to be mapped, as in tomographic reconstruction.

When a source is well away from the meridian the inequality of path differences of waves arriving at the different antennas would cause the fan beams to be smeared out because of the substantial bandwidth of 60 MHz that is utilized to increase sensitivity. To compensate, variable delay lines of effective length up to 100 m are incorporated in the transmission lines leading the received signals to a central system of ten multipliers that generate the complex correlation between the fields received at the ten possible spacings between pairs of elements. Fourier transformation of the correlation brings the data into the space domain in preparation for map reconstruction.

The irregular spacing of the 5 antennas provides for all 9 spacings from D up to $9D$ without incurring the cost of providing a uniformly spaced array of 10 elements. The autocorrelation of the configuration function will show how the chosen arrangement minimizes the redundancy associated with the uniformly spaced array of the same overall length.

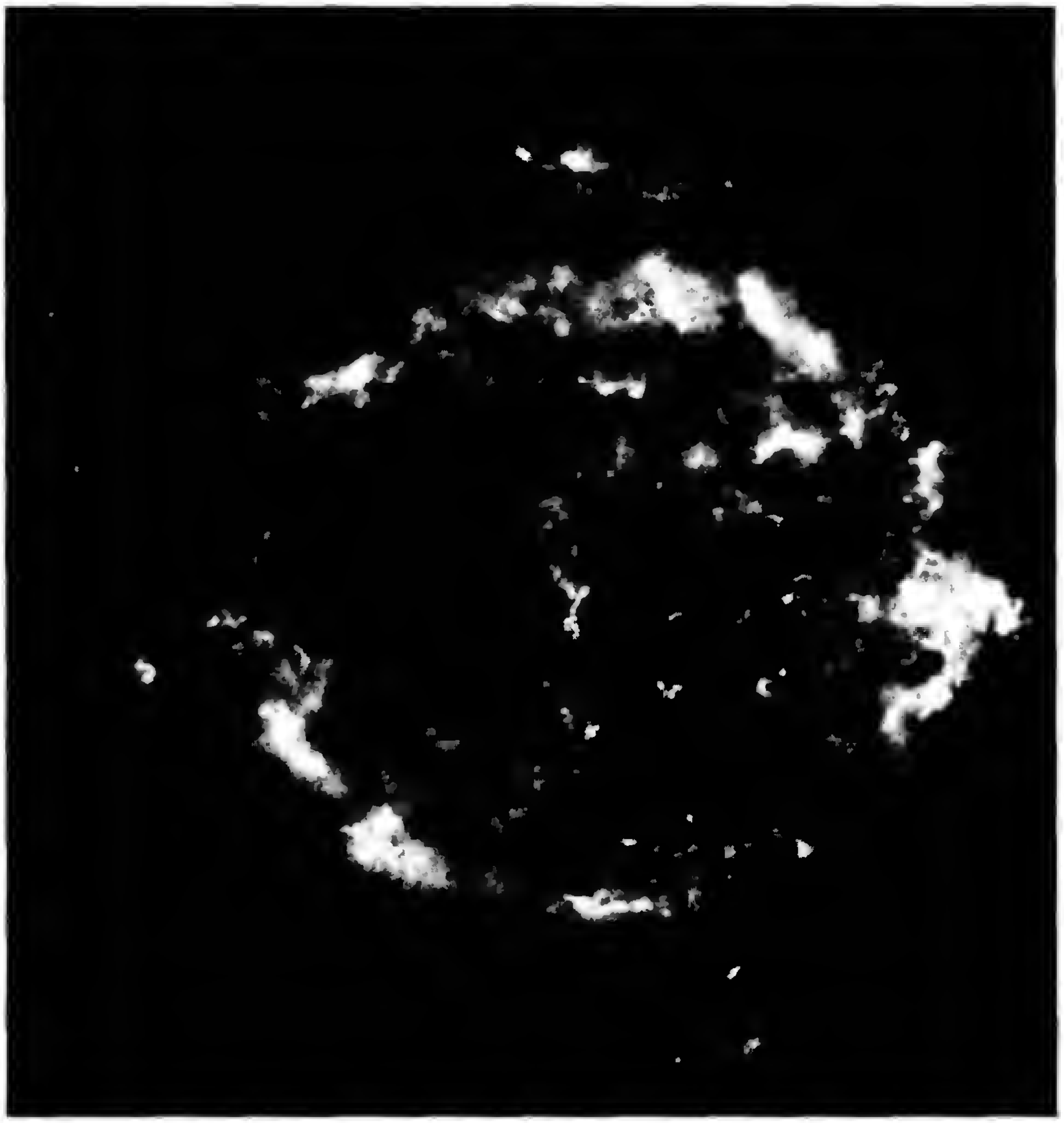


Photo C A Radio Astronomical Image.

A triumph of Fourier synthesis, this image of Cassiopeia A, one of the first radio "point" sources to be discovered in the 1940s, shows the angular diameter to be just under 6 arcminutes. The observations were carried out with the Very Large Array of the National Radio Astronomy Observatory at Socorro, New Mexico. The 27 movable 25-m antennas are arranged in a variety of Y configurations extending up to 21 km in order to measure the two-dimensional complex correlation function that is the Fourier transform of the desired brightness distribution. At a resolution of 0.2 arcseconds and very high dynamic range, the quality of the radio "image" cannot be matched by optical reproduction on film.

The object is the remnant of a supernova explosion that occurred in AD 1680. Its angular diameter has been (measurably) increasing, but as the expanding shell is slowed down by interstellar gas and dust its expansion has slowed. Consequently, interior material that was not in the forefront of the explosion has been able to overtake the shell and punch through, leaving crater-like pits over the surface.

Photo D Tomogram of the Human Head.

This striking image of a thin sagittal slice through a human head in the (x,y)-plane was acquired, without the use of surgery, in the Medical Resonance Imaging (MRI) group of the Electrical Engineering Department at Stanford University and kindly provided by Dr. John Pauly.

Grey levels correspond to the density of hydrogen atoms in the various tissues. Thus air spaces appear black, since they contain no water; bones and teeth appear black for the same reason. Fat, on the other hand, composed of hydrogen-rich molecules, appears bright, as seen surrounding the skull, which is black, while muscle and brain show various grey levels according to their content of water and organic molecules. The thin layer of skin overlying the subcutaneous fat can be seen faintly. The conspicuous crinkles in the fat are due to contact with a pillow.

The protons of the hydrogen atoms emit the signals from which the image is constructed. In the presence of a steady magnetic field B_0, a proton, if its spin axis is perturbed, tends to recover by precessing at a frequency γB_0; for example, in a field of one tesla, the precession frequency is 42.58 MHz. The factor γ is the gyromagnetic ratio of the proton. In the course of precessing in the one-tesla field the proton loses energy to radio-frequency emission at 42.58 MHz that is readily detectable by a radio receiver. Perturbation of the proton spin in the presence of a uniform field B_0 is produced by irradiation with the frequency γB_0. All the protons in the object to be imaged will respond but attention can be confined to the content of a single thin slice centered on say $z = 0$ if the imposed magnetic field has a gradient G_z measured in $T\mathrm{m}^{-1}$ in the z-direction. Then the emission from any one slice perpendicular to the z-axis can be selected with a narrow-band radio receiver tuned to $\gamma(B_0 + G_z z)$. The exciting pulse must now be brief enough to spread over the required frequency range.

Let the complex time-varying pulse representing the total radiation in the frequency band $\gamma(B_0 + G_x x)$ corresponding to the volume of interest be $s(t)$. Then if the density of protons in the (x,y)-plane is $M(x,y)$,

$$s(t) = \int\int M(x,y)e^{-i2\pi\gamma G_x tx}dxdy,$$

which is the value of the Fourier transform of $M(x,y)$ at frequency $\gamma G_x t$. If we allow G_x to vary with time, and add another linearly varying field G_y for the y-axis, the transform can be traced out. This example used a raster scan, where 256 separate lines of the transform were acquired sequentially, but other trajectories in transform space, such as helices, spirals or spokes, defined by varying the gradients appropriately, are also used, especially where volumetric imaging for subsequent display plane by plane is practiced.

15

Antennas and Optics

Antenna theory is sufficiently analogous to the theory of waveforms and spectra that familiarity with one field can be carried over to the other, once one has the key. The first step is to establish a Fourier transform relation in this subject and then to see what the physical quantities are that enter into it. Since our purpose is to improve the versatility of the Fourier transform as a tool, only one aspect of antenna theory is touched on, but it is of fundamental and far-reaching importance in thinking about antennas.

We approach the subject by considering one-dimensional apertures. This may seem inappropriate, because the essence of antenna theory is that a higher dimensionality is required than suffices for the study of waveforms, which are functions of only one independent variable. The one-dimensional treatment is adequate, however, for discussing a great many antennas whose directivity is separable into a product of directivities of one-dimensional apertures, and it is sufficient for setting up the analogy. Features of the two-dimensional theory are introduced later.

Much of the material of this chapter is directly applicable, with changes in terminology, at optical as well as at radio wavelengths; thus the parallel light incident on an aperture is diffracted into space in much the same way as radio emission from an aperture which is of the same shape (but larger, in proportion to its wavelength). In one class of problem an irradiating field, optical or radio, is given, but the aperture illumination is not specified explicitly. It is often sufficient to assume that the illumination that an incident wave produces in an aperture in a screen is the same as would have been produced if the screen had not been there, but since these more subtle questions of diffraction theory are not to be considered here, our point of departure is the aperture distribution itself. In radio the excitation of an aperture, both in amplitude and in phase, may be directly imposed by separate transmission lines leading to an array of closely spaced dipoles, a possibility that does not exist at optical wavelengths. For this reason it seems

appropriate for a discussion of diffraction from given aperture distributions to be couched in terms of antennas.

ONE-DIMENSIONAL APERTURES

An electromagnetic horn feeding energy into semi-infinite space, as shown in Fig. 15.1, is an example of a type of antenna which can conveniently be replaced by a plane-aperture model. As far as the flow of energy into the right-hand half-plane is concerned, we may say that the source is equivalent to a certain distribution of electric field over an aperture plane AA'. The field is zero over the part of the plane occupied by conductor, but some definite distribution is maintained over the horn opening by the source of electromagnetic energy.

Suppose that all conditions are independent of y, the coordinate perpendicular to the plane of the paper. Then the antenna will be called a one-dimensional aperture. It is also supposed in most of what follows that the opening is at least several wavelengths across. This proves to be equivalent to assuming antennas of high directivity.

Since any antenna may be replaced, as far as its distant effects are concerned, by a plane distribution of field, restriction to plane apertures is not in principle destructive of generality. But some antennas (particularly end-fire arrays such as the Yagi antenna) are so nonplanar that they are normally considered by other methods.

Two-dimensional apertures are readily handled as a generalization of the one-dimensional case.

Let the aperture distribution of electric field at the point x at time t be $F(x) \cos [\omega t - \phi(x)]$, where ω is the angular frequency of the (monochromatic)

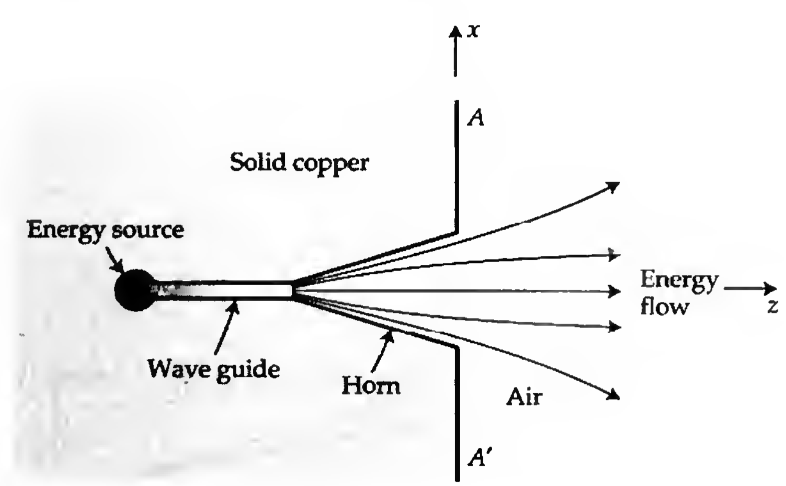


Fig. 15.1 An antenna which can be regarded as a plane aperture.

oscillation and $F(x)$ and $\phi(x)$ are the amplitude and phase, respectively. The phasor $E(x)$, that is, the complex quantity given by

$$E(x) = F(x)e^{-i\phi(x)},$$

will be understood wherever electric fields are dealt with below. We note that the instantaneous field is given by the real part of $E(x)$ exp $i\omega t$; that is,

$$F(x)\cos[\omega t - \phi(x)] = \mathrm{Re}\,[E(x)e^{i\omega t}].$$

By Huyghens' principle, the electric field produced at a distant point will be the sum of the effects of the elements of the aperture. Each element produces an effect proportional in amplitude to the field at the element and retarded in phase by the number of cycles contained in the path to the distant point. Thus at a distance r the field due to the element between x and $x + dx$ is proportional to

$$E(x)\,dxe^{-i2\pi r/\lambda},$$

where λ is the wavelength.

Consider the field at a distant point P in a direction making an angle θ with the z axis (Fig. 15.2). Let the point P be at a distance R from the origin $x = 0, z = 0$. Then to an approximation,

$$r = R + x\sin\theta$$
$$= R + xs$$

where $s = \sin\theta.$

Then $E(x)e^{-i2\pi r/\lambda}dx$

becomes $E(x)e^{-i2\pi R/\lambda}e^{-i2\pi xs/\lambda}dx.$

Let the factor $\exp(-i2\pi R/\lambda)$, which expresses the average phase retardation of the effects of the aperture elements at a distance R, be lumped with the complex constant of proportionality. Then, integrating all the separate effects, we find that the distant field, in the direction s, is given by

$$\int_{-\infty}^{\infty} E(x)e^{-i2\pi xs/\lambda}\,dx.$$

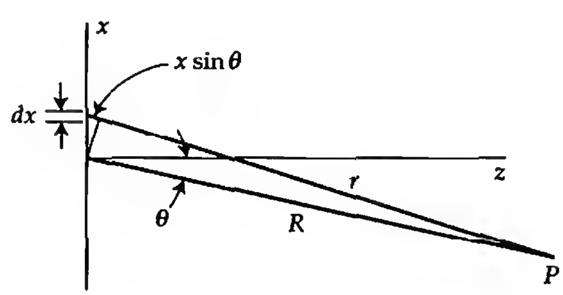


Fig. 15.2 Construction for applying Huyghens' principle.

The distant field, as a function of direction, is thus seen to be related to the aperture field distribution through the Fourier transformation. To throw the basic relation into the standard form, we henceforth measure distances in the aperture plane in terms of the wavelength; that is, we deal in x/λ instead of x. We then define $E(x/\lambda)$ to be equal to the old $E(x)$ and introduce $P(s)$ such that

$$P(s) = \int_{-\infty}^{\infty} E\left(\frac{x}{\lambda}\right) e^{-i2\pi(x/\lambda)s} d\left(\frac{x}{\lambda}\right).$$

The quantity $P(s)$, which is proportional to the distant field produced in the direction s, is known as the field radiation pattern or angular spectrum. Usually it is normalized in some way, but as we have introduced it, it depends not only on the character of the aperture excitation $E(x)$ but also on the level of excitation. When normalized, it is characteristic of the antenna alone, irrespective of the strength of the energy source.

By the reciprocal property of the Fourier transformation, it then follows that

$$E\left(\frac{x}{\lambda}\right) = \int_{-\infty}^{\infty} P(s) e^{i2\pi(x/\lambda)s} ds.$$

Since s is the sine of an angle, the meaning of an integral taken between infinite limits of s may be obscure. However, the restriction to high directivity ensures that the integrand falls to zero while s is still well within the range -1 to $+1$. The precise limits then do not matter, provided only that they are large enough.

As an example consider an aperture for which

$$E(x_\lambda) = \Pi\left(\frac{x_\lambda}{w_\lambda}\right),$$

where $x_\lambda \triangleq x/\lambda$;
that is, a uniform field is maintained over a distance of w_λ wavelengths, outside which the field is zero. Then

$$P(s) = w_\lambda \operatorname{sinc} w_\lambda s$$

as shown in Fig. 15.3.

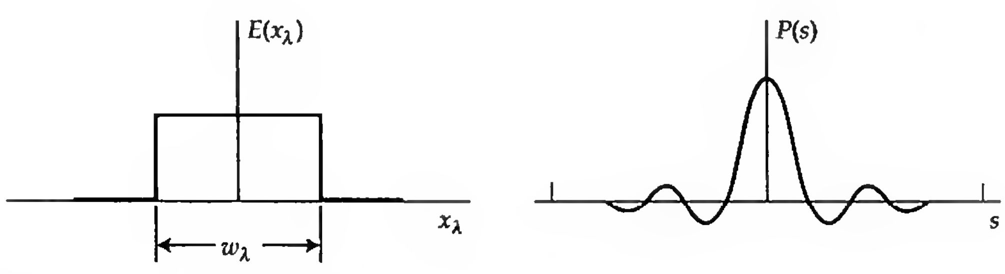


Fig. 15.3 An aperture distribution and its angular spectrum.

In the one-dimensional theory that follows, we shall restrict the discussion to highly directive antennas. In the more general case the following phenomena enter the discussion. First, the polarization of the field must be taken into account; in directions around the normal to the aperture, the radiated energy is approximately the same whether the aperture field is in the x or the y direction, but at wide angles a cosine factor enters in one case and not in the other. Another factor allows for the difference which sets in between the radian and the unit of s. Then sharp edges in the aperture distribution must be allowed for; they set up evanescent fields which die away with increasing z and do not contribute to the radiated energy. In a highly directive antenna the energy of the evanescent fields is negligible in comparison with the radiation.

ANALOGY WITH WAVEFORMS AND SPECTRA

Having established a Fourier transform relation between a pair of physical quantities in the subject of antenna theory, we are in a position to translate statements about antennas into statements about waveforms and spectra, and vice versa. It only remains to set up a table of corresponding terms as follows:

Waveforms and spectra	Antennas
Time t	Aperture distance in wavelengths x_λ
Frequency f	Direction sine s
Waveform $V(t)$	Aperture field distribution $E(x_\lambda)$
Spectrum $S(f)$	Field radiation pattern or angular spectrum $P(s)$

In view of the reciprocity of the Fourier transformation, it is possible to set up the correspondence in two different ways. The one given here is customary, as evidenced by terminology such as "angular spectrum." However, there are two separate and distinct antenna analogues to any one waveform-spectrum pair, and we draw on both as needed.

It will be noticed that $P(s)$ is the minus-i transform of $E(x_\lambda)$, just as $S(f)$ is the minus-i transform of $V(t)$. Clearly this is a convenience in the present context, but the opposite choice might just as well have been made. For the present it is desirable to bear in mind that x and θ are measured in opposite directions.

If we know that a rectangular pulse has a spectrum which is a sinc function of frequency, it follows that a uniformly excited finite aperture produces an angular spectrum of radiation which is a sinc function of direction.

Knowing that an impulse has a flat spectrum tells us that a very small aperture radiates isotropically, and the fact that an impulse pair has a cosinusoidal spectrum means that two infinitesimal antennas produce a cosinusoidal variation of field on a distant plane. These cases are illustrated in Fig. 15.4, and any of the Fourier transform pairs which have been encountered may be so interpreted.

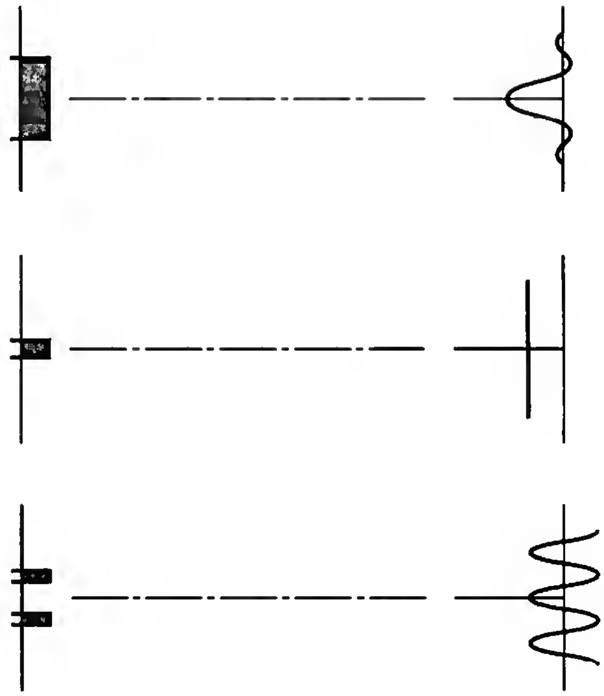


Fig. 15.4 **Some aperture distributions and the field that they produce on a distant plane.**

In addition to this possibility of reinterpreting particular transform pairs, there are other classes of information which may be transferred from our body of knowledge about waveforms and their spectra. For example, all the theorems given earlier may be endowed with new meaning in the new physical field. Furthermore, all the corresponding properties investigated previously may be translated into corresponding properties of aperture distributions and their angular spectra. As a first example, consider the similarity theorem.

BEAM WIDTH AND APERTURE WIDTH

The similarity theorem tells us that if the dimensions of an aperture are scaled up, the form of the aperture distribution being kept the same, then the field radiation pattern retains its form and is compressed by the scaling factor, for if

$$E(x_\lambda) \quad \text{gives rise to} \quad P(s),$$

then

$$E(ax_\lambda) \quad \text{gives rise to} \quad |a|^{-1}P\left(\frac{s}{a}\right).$$

Thus the wider the aperture, the narrower the beam.

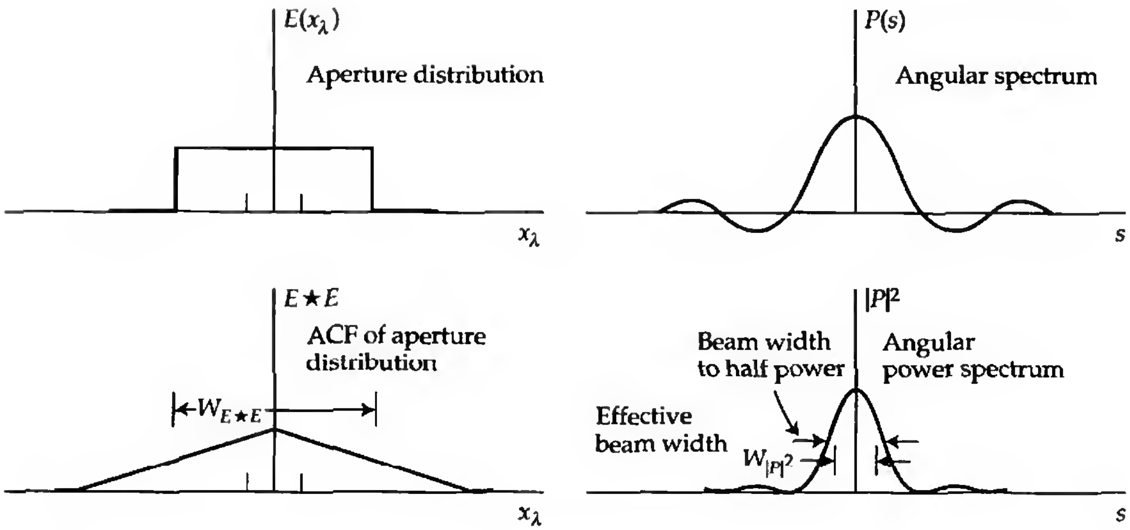


Fig. 15.5 **Relations between some antenna quantities.**

The quantity $|P(s)|^2$ will be referred to as the angular power spectrum. Its Fourier transform is the autocorrelation function of the aperture distribution. From the reciprocal relation between the equivalent widths of a function and its transform it follows that, given an aperture distribution, we can calculate the effective beam width of its radiation by taking the reciprocal of the equivalent width of the autocorrelation function of the aperture distribution. For example, if we have a uniform aperture 10 wavelengths wide, the equivalent width of the (triangular) autocorrelation function will be 10 wavelengths; hence the equivalent width of the angular power spectrum will be one-tenth of a radian, or 5.7 degrees. This example is illustrated in Fig. 15.5.

BEAM SWINGING

According to the shift theorem, if a waveform is shifted with respect to $t = 0$, the result is a progressive phase delay increasing with increasing frequency f. One must therefore be able to shift the beam of an antenna (with respect to $s = 0$) by introducing a linear gradient of phase along the aperture coordinate x. In fact the shift theorem states that

$$E(x_\lambda)e^{-i2\pi S x_\lambda} \supset P(s+S).$$

Hence, to shift the beam by an amount S it is necessary to introduce a phase gradient of S cycles per wavelength along the aperture.

Various methods of beam swinging are in use which depend on this principle. A direct approach is to insert phase shifters or to introduce extra lengths of

transmission line. Extra air paths may be introduced, as in the method of swinging the beam of a parabolic reflector by displacing the feed point from the focus. Small shifts in frequency can also be used to introduce a suitable phase gradient. If a thin dielectric prism were placed on the aperture, it would introduce a progressive phase delay depending linearly on x, and the beam would consequently be shifted. The amount of shift would be precisely that which we would calculate on the basis of the prismatic refraction.

ARRAYS OF ARRAYS

In its application to antennas, the convolution theorem parades as the array-of-arrays rule. According to this well-known rule, the field pattern of a set of identical antennas is the product of the pattern of a single antenna and an "array factor." This factor is the pattern which would be obtained if the set of antennas were replaced by a set of point sources.

Let the aperture distribution of a certain antenna be $E(x/\lambda)$. Then the distribution

$$\sum_{n=1}^{N} E\left(\frac{x - x_n}{\lambda}\right)$$

is a set of N such antennas centered at the points $x = x_1, x_2, \ldots, x_N$. Such a distribution may be expressed as a convolution

$$E\left(\frac{x}{\lambda}\right) * q\left(\frac{x}{\lambda}\right),$$

where $q(x/\lambda)$ is a set of impulses,

$$q\left(\frac{x}{\lambda}\right) = \sum_{n} \delta\left(\frac{x - x_n}{\lambda}\right).$$

Let $Q(s)$ be the Fourier transform of $q(x/\lambda)$. It follows from the convolution theorem that if a single antenna has a pattern $P(s)$, then the set of antennas has a pattern

$$P(s)Q(s).$$

Here $Q(s)$ is the array factor.

We may also apply the convolution theorem in reverse. If we begin with an aperture distribution $E(x/\lambda)$ and then multiply it by some factor depending on x, then the effect on the beam may be deduced by convolving the original beam with the transform of the factor. For example, many aperture distributions are approximately uniform, departing from uniformity only in that the excitation falls off toward the edges. This broad tapering factor has a narrow transform whose convolution with the pattern of the uniform distribution makes the necessary

modification. The multiplying factor need not be real; suppose that it is exp ikx, that is, a factor whose modulus is unity and whose phase is proportional to x. Its transform is a displaced impulse. Convolution with a displaced impulse simply displaces the beam.

INTERFEROMETERS

The modulation theorem, a special case of the convolution theorem, has particular significance for antennas. Consider a pair of identical antennas spaced well apart. Such an arrangement is called a two-element interferometer. Since the full distribution can be expressed as a convolution between a single element and a pulse pair of the appropriate spacing, the pattern may be obtained from that of a single antenna simply by multiplying it by a suitable cosine function of s.

Thus Fig. 15.6 shows an aperture distribution

$$E(x_\lambda) = \Pi\left(\frac{x_\lambda + W_\lambda}{w_\lambda}\right) + \Pi\left(\frac{x_\lambda - W_\lambda}{w_\lambda}\right)$$

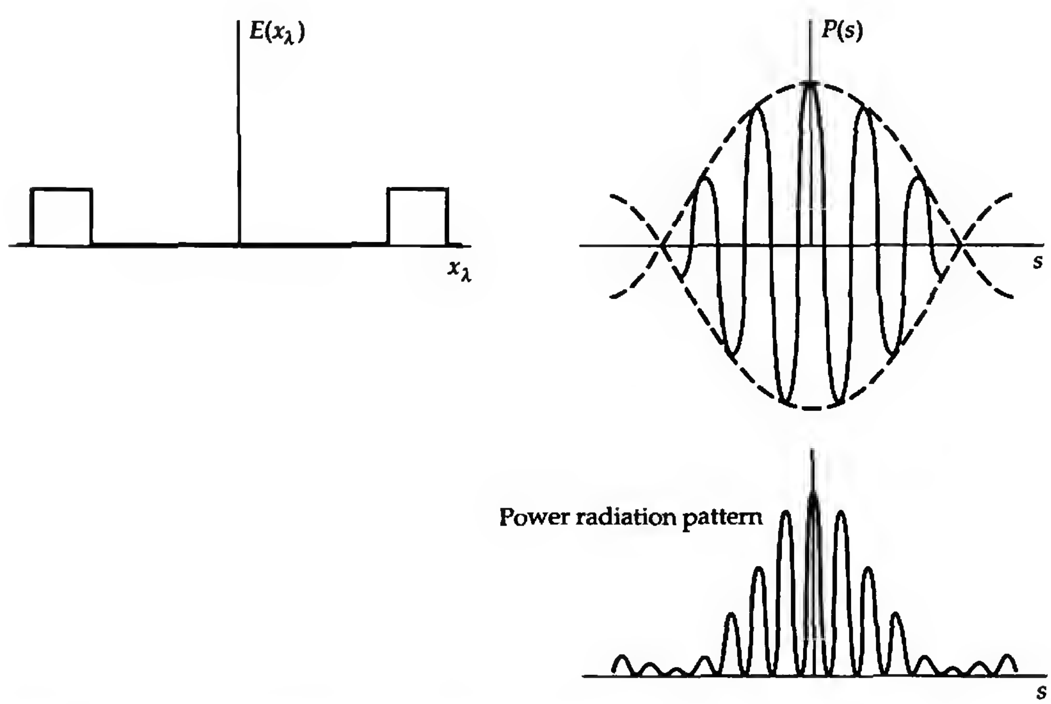


Fig. 15.6 **An aperture distribution, representing a two-element interferometer, its angular spectrum $P(s)$, and its power radiation pattern.**

which could be set up by means of two antennas spaced $2W_\lambda$ connected to a central pair of terminals by equal lengths of cable. It follows that

$$P(s) = 2w_\lambda \operatorname{sinc} w_\lambda s \cos 2\pi W_\lambda s,$$

and that the power radiation pattern is approximately the square of this, as shown in the figure.

SPECTRAL SENSITIVITY FUNCTION

When the first sky maps of radio emission were made it was apparent that the observed distribution of radio intensity as a function of direction could not be exactly the same as the true distribution because of averaging of the power received from different directions within the antenna beam. But because of the spherical curvature of the sky and the wide beams of 10° or more that were then at the limit of antenna practice, the connection with two-dimensional convolution was not immediately apparent. The notion that a celestial brightness distribution could be analyzed into spatially sinusoidal components and that the response of a scanning antenna would be sinusoidal was not obvious, and indeed is only an approximation. However, when a small patch of sky is scanned by a narrow beam the situation is analogous to signal theory. The true distribution of brightness is the input signal; the power radiation pattern of the antenna is the impulse response of a linear system; and the output record is the convolution between the true distribution and the antenna pattern. Therefore, by the convolution theorem, there is a transfer function describing the sensitivity of the system to different spectral components. Spatially uniform sky brightness can now be spoken of as spatial d.c. The response to sinusoidal components of brightness weakens as the spatial frequency u, measured in cycles per radian, increases; the spatial minima, straddled by the beam, are partly filled in by contributions from brighter areas that are adjacent to the minima but still within the beam, while spatial maxima are eroded by inclusion within the beam of regions that are less bright. Consequently, the amplitude of a sinusoidal component, when scanned by an antenna, is reduced. As a function of spatial frequency u, the reduction factor, relative to spatial d.c., is the spectral sensitivity function. From the convolution theorem it follows that the autocorrelation function of the antenna aperture distribution (see Fig. 15.5) expresses the dependence of sensitivity on spatial frequency. For example, for an aperture distribution $E(x/\lambda) = \Pi(x/w)$ as in Fig. 15.3, the spectral sensitivity function $T(u)$ equals $\Lambda(wu/\lambda)$, normalized to $T(0) = 1$. We learn from this that a simple antenna of width w responds to different spatial frequencies in a calculable way, that there is a critical spatial frequency $s_c = \lambda/w$ above which the antenna does not respond at all, and that this cutoff is the same for all antennas representable by an aperture distribution of width w.

Thus the analogy with signal filters has a curious twist; a filter with impulse response of finite duration has a frequency response that runs on, while the analogous frequency response of an antenna cuts off. The ability to translate antenna mapping terminology into terms of filter theory is discussed by Bracewell and Roberts (1954) and Bracewell (1958, 1962, 1995).

MODULATION TRANSFER FUNCTION

An image created by a camera lens depends on the scene in a way that is related to antenna mapping as follows. First the three-dimensional scene is imagined to be projected onto an object plane parallel to the film plane in the camera. Then the light from each point of the object will be smeared into a spot of light on the film described as the point response function of the lens. Despite the effort of lens design, the point response function varies over the image plane, but is taken to be invariant for practical purposes over limited regions referred to as isoplanatic patches. Within such a patch a simple convolution relation holds between object and image. If the object brightness is a spatial sinusoid then the image will be sinusoidal too, but of reduced amplitude as given by a visibility function that is normalized to be unity for spatial d.c. or uniform brightness. There may in general be a spatial phase shift although for the central isoplanatic patch the circular symmetry of a lens makes phase shift unimportant.

The visibility function played a basic role in the measurement of stellar diameters by A. A. Michelson (1902) in observations made visually of the amplitude of sinusoidal diffraction fringes seen in a telescope when the aperture was blocked except for two parallel slits. By noting how the visibility of the fringes fell off with increased slit spacing. Michelson obtained part of the Fourier transform of the brightness distribution over Betelgeuse and deduced an angular size parameter. Atmospheric variation causes the diffraction fringes to be in motion in the eyepiece; consequently spatial phase was ignored. The descriptive term fringe visibility in optics came to be replaced by modulation transfer function (MTF); it corresponds to the absolute value of the spectral sensitivity function. When Fourier methods were introduced into lens design the modulation transfer function became an important tool for characterizing the performance of a lens in more detail than the customary resolution expressed in lines per cm, a single parameter easily obtainable by photographing ruled grids.

When spatial phase came to be taken into account the term optical transfer function (OTF) was adopted for the complex Fourier transform of the point response function; its absolute value is the old modulation transfer function. From the point of view of radio interferometry, fringe visibility generalizes into complex fringe visibility, which takes into account not only the observed fringe amplitude but also any spatial displacement expressed in terms of the fringe spacing (Bracewell, 1958). In electronics, the complex factor $T(f)$ (Chapter 9) relating the output sinusoid to the input sinusoid, as well as $|T(f)|$, may both be called the transfer function.

PHYSICAL ASPECTS OF THE ANGULAR SPECTRUM

Fourier analysis of an aperture distribution $E(x/\lambda)$ is in terms of components of the form

$$P(s)e^{i2\pi x_\lambda s}.$$

Such a component is one whose amplitude is the same at all points x of the aperture but whose phase advances progressively along the aperture. It is the field that would be set up by a plane wave incident on the aperture from behind at a suitable angle to account for the phase gradient. (The angle of incidence is one whose sine is s.) Naturally the aperture field set up by a certain incident plane wave is the aperture field which will launch a plane wave in the same direction as the incident one. Thus each component of the aperture distribution launches a plane wave in a certain direction, and the total radiation from the aperture in all directions is expressible as a combination of infinite plane waves in all directions, each with suitable amplitude and phase.

When $|s| > 1$, the phase $2\pi x_\lambda s$ at the point x has a rate of change exceeding one full cycle in a distance λ. It is apparent that no incident traveling wave can set up such a field because the crest-to-crest distance measured along any oblique section of the wave can exceed or equal the wavelength λ, but cannot be less. If such an aperture field is imposed, no traveling wave will be launched; the field will simply die away in the z direction.

Instead of analyzing into exponentials, we can break the aperture distribution down into cosine and sine distributions of the form

$$\begin{smallmatrix}\cos\\ \sin\end{smallmatrix}(2\pi x_\lambda s).$$

The spatial period of such a component, measured in units of x, is λ/s, or the "spatial frequency" is s/λ cycles per unit of x. Since each such component is composed of a pair of exponentials, it launches a pair of equal plane waves in directions equally inclined to the x direction. The lower the spatial frequency, the closer these two waves are to the z direction, and conversely, as the spatial frequency approaches one cycle per freespace wavelength λ, we have the standing wave set up by two waves traveling in opposite directions in the plane of the aperture.

This often-helpful approach to the angular distribution of the radiation from an aperture is an alternative to the treatment by Huyghens' principle, where the aperture is analyzed into elements.

TWO-DIMENSIONAL THEORY

Since the radiation from an aperture distribution essentially involves two-dimensional transform theory, the one-dimensional presentation should be re-

garded as an introduction to the subject. The following discussion sets up the Fourier transform relation in the two-dimensional case and without the restriction referred to above as the high-directivity approximation.

Let the distribution of electric field over an aperture plane (x,y) have a component $E_y(x_\lambda,y_\lambda)$ in the y direction, and let the x component be zero. If it is not, it may be handled separately in the same way as the y component. There will be a z component normal to the aperture plane, but since it can be deduced from the y component by applying the condition that the divergence of the electric field vector should be zero, it is not needed in a specification of the aperture distribution.

The field set up by the aperture can be expressed as a sum of plane waves of the form

$$\mathcal{P}(l,m)e^{i(2\pi/\lambda)(lx+my+nz)}\,dl\,dm,$$

where l,m,n are direction cosines and $\mathcal{P}(l,m)\,dl\,dm$ is the complex amplitude of the waves in the cone l to $l + dl$, m to $m + dm$.

A wave traveling in the direction (l,m,n) produces on the $z = 0$ plane a field

$$\mathcal{P}(l,m)e^{i(2\pi/\lambda)(lx+my)}$$

whose spatial period λ' is given by

$$\lambda' = \frac{\lambda}{(l^2 + m^2)^{\frac{1}{2}}}.$$

If $l^2 + m^2$ exceeds unity, the spatial period λ' is less than λ and thus represents fine detail in the aperture distribution. In this case n is imaginary, because of the property of direction cosines that

$$l^2 + m^2 + n^2 = 1,$$

and a component plane wave takes the form

$$e^{-(2\pi/\lambda)|n|z}\mathcal{P}(l,m)e^{i(2\pi/\lambda)(lx+my)}\,dl\,dm,$$

which represents a wave traveling in the (x,y)-plane, but decaying exponentially with z. Such a field does not contribute to the radiation field at any great distance from the aperture.

If $l^2 + m^2$ is less than unity, λ' ranges from infinity (propagation normal to aperture plane) to λ (propagation parallel to plane). The power radiated per unit solid angle in the direction (l,m) is proportional to the squared modulus of the complex amplitude of the waves proceeding, per unit solid angle, in a small cone of directions centered on (l,m). Since the solid angle subtended by an element $dl\,dm$ is $n^{-1}\,dl\,dm$, the radiation pattern is

$$\left|\frac{\mathcal{P}(l,m)\,dl\,dm}{n^{-1}\,dl\,dm}\right|^2 = n^2|\mathcal{P}(l,m)|^2.$$

To find the aperture distribution $E_y(x_\lambda,y_\lambda)$ associated with a given $\mathcal{P}(l,m)$, we sum the component plane waves on $z = 0$ for all l and m. The cosine of the angle β between the field of the ray (l,m) and the y axis is $n(1 - l^2)^{-\frac{1}{2}}$. Hence, resolving the field of each ray into the y direction we have

$$E_y(x_\lambda,y_\lambda) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{n}{(1 - l^2)^{\frac{1}{2}}}\mathcal{P}(l,m)e^{i(2\pi/\lambda)(lx+my)}\,dl\,dm.$$

This is not quite the classical Fourier transformation. Let the Fourier transform of $E_y(x_\lambda,y_\lambda)$ be $P(l,m)$; it will be referred to as the angular spectrum produced by the aperture distribution. Thus

$$P(l,m) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}E_y(x_\lambda,y_\lambda)e^{-i(2\pi/\lambda)(lx+my)}\,dx_\lambda dy_\lambda,$$

and conversely,

$$E_y(x_\lambda,y_\lambda) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}P(l,m)e^{i(2\pi/\lambda)(lx+my)}\,dl\,dm.$$

It follows that the relation between $P(l,m)$ and the complex amplitude $\mathcal{P}(l,m)$ is

$$\frac{n\mathcal{P}(l,m)}{(1 - l^2)^{\frac{1}{2}}} = P(l,m).$$

Here the radiation pattern is given by

$$\begin{aligned} n^2|\mathcal{P}(l,m)|^2 &= (1 - l^2)|P(l,m)|^2 \\ &= \cos^2\phi|P(l,m)|^2, \end{aligned}$$

where ϕ is the angle between the direction (l,m) and the yz plane. In the case of highly directional antennas radiating in directions close to the z direction, $\cos\phi$ is approximately equal to unity, and $|P|^2$ is essentially the same as the radiation pattern.

OPTICAL DIFFRACTION

Since the electromagnetic theory governing optical radiation is the same as for antennas, there is a fundamental analogy between optics and antenna theory. Indeed there is no clear boundary between radio waves and optics, as witnessed by the field of microwave optics. However, most optical devices are very much smaller than antennas, which leads to conspicuous differences in fabrication and use, even if there is a certain range of overlap. With the passage of time optical

excitation by optical fibers, or by microcircuits that are recognizably derived from microwave circuitry, have blurred the distinction between antennas and optics in practice, and it has become common to combine the fundamentals of both for some teaching purposes.

Quantum effects are not significant for antennas except at the very highest frequencies or at very low temperatures. A more striking difference between a telescope and a receiving antenna is that the telescope captures an image in its focal plane while an antenna usually does not. There is indeed an image in the focal plane of a paraboloidal reflector, but only one pixel is normally captured, for example by a single electromagnetic horn centered on the focal point. It is as though the film or CCD array in the focal plane of a telescope was replaced by a single photodetector. So when point response function is mentioned, in optics that is an entity that is recorded in its entirety by an exposure on a photographic plate or film, whereas for an antenna a two-dimensional raster scan has to be performed. A closer analogy between an antenna and a telescope is provided in the practice of precision photometry of stars, where telescopes are in use that are provided with a single photodetector in the focal plane.

FRESNEL DIFFRACTION

In the discussion of antennas the objective was to relate a plane aperture distribution of field to the radiation pattern as a function of direction in the distant field. Appropriately for this book the development was based on the Fourier transform of the aperture distribution, or angular spectrum. Because of the uniqueness relationship between a function and its Fourier transform, the angular spectrum contains all the information that is contained in the specification of the aperture distribution, and thus provides a rigorous basis. However, with antennas, the distant radiation pattern is of more importance to communication links and received the emphasis.

Where does the distant field set in? If the aperture is N wavelengths across, the distant field, far field, or Fraunhofer region, sets in at a distance of N aperture widths. Communication ordinarily takes place over distances much greater than this. There are exceptional cases at very low frequencies, for example with radars studying the ionosphere, where the near field extends to several tens of kilometers; also, the elements of many antenna arrays are close enough together to be in each other's near fields, which becomes a factor in the antenna design. In optics on the other hand the near field, or Fresnel field, is of everyday importance. Conveniently, the Fresnel field may also be discussed by considering the Fourier components of the aperture distribution.

As before, let the aperture plane, or function domain, be the (x,y)-plane, while the z-axis is normal to the aperture plane. A Fourier component of the aperture distribution is a spatially stationary sinusoid such as $\cos k_x x$, where k_x is the angular spatial frequency in radians per unit of x (radians per meter for example).

The crest lines of this particular wave are parallel to the y-axis, while the z-axis passes through a crest line. The general component is of the form

$$a \cos(k_x x + k_y y) + b \sin(k_x x + k_y y),$$

which allows for any orientation on the (x,y)-plane and any position of the z-axis with respect to a crest line. But, by rotating and shifting the (x,y)-axes we can return to $\cos k_x x$ with whatever amplitude we please, and this expression will suffice for the following discussion.

As we have seen, such a standing wave may be created by shining onto the (x,y)-plane, two equal plane waves of wavelength λ described by rays equally inclined at angle i to the z-axis and coplanar with it. It follows that the standing wave $\cos k_x x$, however created, will radiate a pair of equally inclined rays. The angle of incidence will be related to k_x by $\sin i = k_x \lambda / 2\pi$. An alternative parameter λ_x, the wavelength in the x-direction, is related to k_x by $k_x = 2\pi/\lambda_x$. When the two traveling waves are added together at $z = 0$ the result is the standing wave $\cos k_x x$, or $\cos(2\pi x/\lambda_x)$. But away from the (x,y)-plane, where $z > 0$, the constituent traveling waves will combine, with delayed phase, to produce a reduced copy given by $\cos k_x x \cos k_z z$, where $k_x^2 + k_z^2 = (2\pi/\lambda)^2$. At a distance $z = 2\pi/k_z = \lambda_z$ from the aperture plane, the field will have returned to $\cos k_x x$. The field thus propagates in the z-direction through a distance λ_z in one cycle. Since λ_z must be greater than λ (provided $i \neq 0$), the speed in the z-direction is λ_z/λ times the speed of the incident waves. This is exactly the situation in a waveguide that extends in the z-direction and is enclosed by conducting walls at $x = \pm \lambda_x/4$.

To understand this configuration better, consider two extreme cases. When $i = 0$, both incident waves coincide, λ_x is infinite, $\cos k_x x$ becomes unity independent of x, and there is uniform illumination over the aperture plane. Naturally this aperture distribution launches a single ray in the z-direction. If i is just a little less than $\pi/2$, and increasing, then the incident waves pass through each other in almost exactly opposite directions and produce a resultant wave in which $\lambda_x \approx \lambda$, while $\lambda_z \to \infty$. This might seem to exhaust the possible cases but it is perfectly possible to create an aperture distribution where $\lambda_x < \lambda$. This can be done by inserting an iris diaphragm in a waveguide or by oblique illumination as with total internal reflection inside glass. As z increases the field then decays exponentially, or evanesces, as $\cos k_x x \exp(-|k_z|z)$.

From this physical picture one can construct the Fresnel field at any distance z from a given aperture distribution as follows. Fourier analyze the aperture distribution into sinusoids, of various spatial frequencies k_x, multiply each by the corresponding propagation factor $\cos k_z z$, where $k_z = \sqrt{(2\pi/\lambda)^2 - k_x^2}$ and recombine to synthesize the Fresnel field by addition. The presence of evanescent waves, which are not common in everyday applications, will be revealed by imaginary values of k_z; for such components use the corresponding decay factor $\exp(-|k_z|)$.

This procedure is rigorous and is similar to calculating the response of a filter to an input waveform by analyzing into components, altering each separately by a transfer factor, and recombining. It is well adapted to numerical computation of fields in the Fresnel region.

OTHER APPLICATIONS OF FOURIER ANALYSIS

Antennas and optical devices are common parts of measuring instruments, communication links, recording equipment, machines, and elastic structures and as such play a role in a variety of fields. Microwave sensing of the environment and photography are examples where convolution relations exist between what is there and what is recorded. Thus a background in Fourier analysis is useful not only for understanding and designing radio and optical sensors themselves but also for their information-gathering functions. As far as the information aspect is concerned the sensor might be regarded as a black box specified by external properties such as beamwidth. Not only is Fourier analysis a tool for electromagnetic design of the box contents but it is also a tool for the user who is not concerned with such details but is engaged in information gathering and processing. The notion that scanning or mapping with a beam in space is analogous to electrical filtering in time, once a novel and unformulated concept, illustrates the remarkable versatility of the Fourier approach.

To further explore the field of antennas requires a sound background in electromagnetic theory, followed by study of a text such as Kraus (1988), written by one of our outstanding teachers. Further detailed references can be found in other texts listed below and in the lists of references in articles in issues of the *IEEE Transactions on Antennas and Propagation*. To pursue the study of optical applications refer to Goodman (1996), to the various publications of the Optical Society of America and SPIE (The International Society for Optical Engineering), and to other texts mentioned in the bibliography.

BIBLIOGRAPHY

Annual Reports of the Lehrstuhl für Optik, Physikalisches Institut, Universität Erlangen-Nürnberg, Germany.

Born, M., and E. Wolf: "Principles of Optics," 6th ed., Pergamon, New York, 1986.

Bracewell, R. N.: "Radio Interferometry of Discrete Sources," *Proc. Inst. Radio Engrs.*, vol. 46, pp. 97–105, 1958.

Bracewell, R. N.: "Radio Astronomy Techniques," *Encyc. of Physics*, vol. 54. S. Flügge, ed., Springer-Verlag, Berlin, 1962.

Bracewell, R. N.: "Two-Dimensional Imaging," Prentice-Hall, Englewood Cliffs, New Jersey, 1995.

Bracewell, R. N., and J. A. Roberts: "Aerial Smoothing in Radio Astronomy," *Aust. J. Phys.*, vol. 7, pp. 615–640, 1954.

Clarke, R. H., and J. Brown: "Diffraction Theory and Antennas," Ellis Horwood, Chichester; Wiley, New York, 1980.

Collin, R. E.: "Antennas and Radiowave Propagation," McGraw-Hill, New York, 1985.

Gaskill, J. D.: "Linear Systems, Transforms, and Optics," Wiley, New York, 1978.

Goodman, J. W.: "Introduction to Fourier Optics," 2d ed., McGraw-Hill, New York, 1996.

Griffiths, J.: "Radio Wave Propagation and Antennas: An Introduction," Prentice-Hall, Englewood Cliffs, New Jersey, 1987.
Haus, H. A.: "Waves and Fields in Optoelectronics," Prentice-Hall, Englewood Cliffs, New Jersey, 1984.
Kraus, J. D.: "Antennas," 2d ed., McGraw-Hill, New York, 1988.
Marathay, A. S.: "Elements of Optical Coherence Theory," Wiley, New York, 1982.
Michelson, A. A.: "Light Waves and Their Uses," University of Chicago Press, 1902.
O'Neill, O.: "Introduction to Statistical Optics," Addison-Wesley, Reading, Massachusetts, 1963.
Papoulis, A.: "Systems and Transforms with Applications in Optics," McGraw-Hill, New York, 1968.
Ramo, S., J. R. Whinnery, and T. Van Duzer: "Fields and Waves in Communication Electronics," 3d ed., Wiley, New York, 1994.
Schelkunoff, S. A., and H. Friis: "Antennas, Theory and Practice," Wiley, New York, 1952.
Silver, S.: "Microwave Antenna Theory and Design," McGraw-Hill, New York, 1949.
Stutzman, W. L., and G. A. Thiele: "Antenna Theory and Design," Wiley, New York, 1981.
Thompson, A. R., J. M. Moran, and G. W. Swenson, Jr.: "Interferometry and Synthesis in Radio Astronomy," 2d ed., Wiley, New York, 1999.
Yariv, A.: "Optical Electronics," Holt, Rinehart and Winston, New York, 1985.

PROBLEMS

1. Explain why the equivalent width of the angular spectrum $P(s)$ might be unsatisfactory as a general-purpose beamwidth parameter. ▷

2. Draw diagrams on the complex plane of $E(x)$ to illustrate the integral

$$\int_{-\infty}^{\infty} E\left(\frac{x}{\lambda}\right) e^{-i2\pi sx/\lambda}\, d\left(\frac{x}{\lambda}\right).$$

Take $E(x/\lambda)$ equal to unity over a certain range and zero beyond. Take values of s such that the path differences between extreme rays is 0, $\lambda/2$, λ, $3\lambda/2$. ▷

3. Explain with aid of a phasor diagram why the nulls in the pattern of a uniformly excited aperture occur in directions such that the distant fields of the end elements of the aperture are in the same phase.

4. **Monopulse antenna.** An aperture is uniformly excited in amplitude, but one half is in antiphase with the other. Draw phasor diagrams to show that the angular spectrum has a null in the direction normal to the aperture and that the maximum halfway out to the next null is $2/\pi$ of what it would be in the absence of the phase reversal.

5. Investigate by phasor diagrams the location of the first nulls, and the strength and location of the principal maximum in the pattern of a uniform aperture of which one half is in quadrature with the other. ▷

6. **Tolerance to decay.** An antenna is fed from an energy source at one end by a transmission line which excites each element with the same amplitude and phase. Because of deterioration of dielectric components, the transmission line becomes lossy, which causes the amplitude of excitation to diminish with distance from the source end. Investigate the effects on (*a*) the direction in which the beam points, and (*b*) the nulls. ▷

7. Consider the aperture distributions

$$\cos^n \pi x \, \Pi(x).$$

What is the effective beam width in each case? ▷

8. By taking an aperture distribution $f(x)\Pi(x)$ such that the amplitude is greater at the edges than at the center, can the effective beam width be made narrower than in the uniform case $f(x) = 1$?

9. Give physical interpretations of the following theorems and relations in terms of antennas: Rayleigh's theorem, definite-integral relation, uncertainty relation.

10. **Focal-plane image.** A radio image of a celestial object that emits radio waves is formed on a surface containing the focus of a paraboloidal reflector; show that it suffices to sample the image at certain discrete intervals. The collector normally found at the focus of a paraboloid has a finite size; is this size connected with the sampling interval? ▷

11. It is desired to form an x-ray image of an extended celestial object suspected of emitting x-rays, but it is difficult to focus the rays because they are not much refracted or reflected. How would you tackle this problem? ▷

12. Given an example of a highly directive antenna for which the power level of the near side lobes of the radiation pattern goes down as $|\theta|^{-3}$ where the angle θ is measured from the center of the beam. Explain why this law cannot persist out to $\theta = \pi/2$, and work out the actual behavior for your example.

13. **Sidelobe level.** The amplitude of excitation of a circular aperture of radius a is proportional to $a^2 - r^2$. Show that the first side lobe is 26 decibels below the main beam. Mention some reasons that this figure might not be attainable in practice; investigate the magnitude of the perturbation to be expected from a feed-supporting tripod that obstructs 5 percent of the aperture.

14. **Visual acuity.** Calculate the beam width of the eye in milliradians, taking the diameter of the pupil to be 4 millimeters.

15. **Radiation from slot.** The y component of tangential electric field in the aperture of a slot antenna is given by

$$E_y = E_0 \Pi\left(\frac{y}{w}\right) \cos\left(\frac{\pi x}{l}\right),$$

where the width w is small compared with the length l. What is the equivalent width of the beam in the principal directions? In what direction is the radiation zero?

16. **Slot array.** An array of eight parallel slots is cut in the metal undersurface of an aircraft. The width and length of the slots are fixed by structural considerations, and the slot spacing is to be chosen so that the gain is a maximum. The structural expert claims that the slots can be tightly packed because the field intensity on the ground vertically below the aircraft is independent of the spacing, by Huyghens' principle. The antenna expert says the field intensity may be the same, but that extra power will be radiated in unwanted directions; he proposes to install them four-fifths of a wavelength apart. His assistant argues by Rayleigh's theorem that the power radiated is the sum of the powers passing through each slot and concludes that tighter packing is warranted, provided the beam width is unimportant. Identify and correct the wrong statements. ▷

17. **Lunar occultation.** A remote point source of meter-wave radiation emerges from behind the moon's limb. Show that the power received by an antenna varies as

$$\left| \int_{-x}^{\infty} e^{i(-\pi y^2/D\lambda)}\, dy \right|^2,$$

where D is the distance to the moon and x is the distance of the observer from the grazing ray. Let $q(x)$ be the derivative of this expression with respect to x. Show that the Fourier transform of $q(x)$ is given by

$$Q(s) = \lambda D e^{i\pi\lambda Ds^2 \operatorname{sgn} s}.$$

Let the derivative of the power received from an extended source be $g(x)$; show that it has the same autocorrelation function as the true source distribution. Since convolving a function with $q(x)$ merely scrambles its Fourier components by dephasing them in proportion to the square of their frequency, convolution with $q(-x)$ should unscramble them. Examine the merits of

$$q(-x) * g(x)$$

as a restoration of the true source distribution.

18. **Wild's array.** A well-known antenna consists of 96 steerable 42-foot paraboloids equally spaced on a circle 1.5 kilometers in radius, and operates at a wavelength of 3.5 meters. Show that the angular spectrum as a function of the direction cosines l and m is

$$\left\{ \sum_{p=1}^{96} \cos\left[2\pi \times 430 \left(l \cos \frac{2\pi p}{96} - m \sin \frac{2\pi p}{96} \right) \right] \right\} P(r)$$

where $r = (l^2 + m^2)^{\frac{1}{2}}$ and $P(r)$ is the angular spectrum of a single element. Note that the superposition of the 96 cosinusoidal corrugations equally spaced in orientation peaks up at the origin and on certain concentric rings at points where the crests and troughs of the corrugations tend to converge. Show directly that the central peak is to a close approximation described by $J_0(2\pi \times 430r)P(r)$. Show that the precise result is

$$[J_0(2\pi \times 430r) + 2J_{96}(2\pi \times 430r) \cos 96\theta + 2J_{192}(2\pi \times 430r) \cos 192\theta + \ldots]P(r)$$

where $\theta = \arctan(m/l)$. ▷

19. **Ocean wave direction finder.** Show how an array of pressure gauges could be constructed on the sea bottom for directional reception of wave energy, and how by suitable processing of the pressure record from each instrument, various directions could be examined. Why is it that the beam of an antenna array, on the contrary, points in only one direction at a time; and why does there not appear to be any problem analogous to

that of correctly phasing antenna elements? The ocean swell off the coast of California in the frequency range 50 to 100 cycles per kilosecond (wavelengths up to about 1 kilometer) comes principally from storms in the Southern Ocean. Could this be confirmed by a coastal array? ▷

20. Calculate the diffraction pattern of an aperture in the form of an equilateral triangle.

21. **High-resolution interferometer.** Eighteen identical antennas are located at $x = 1, 2, 3, \ldots, 15, 16, 16.5, 32.5$, where the unit of x is large compared with the wavelength. The signals from all the antennas are brought to a common point through identical transmission lines and there the voltages from the first 16 are added, those from the remaining two (at $x = 16.5$ and 32.5) are added, and the two resulting voltages are multiplied together. A point source of radiation passes through the beam of this system. Show that the width of the response is approximately equal to the half-power beamwidth of a single long antenna stretching from $x = 1$ to $x = 48$. ▷

22. **Celestial raster scanning.** The response of a radiometer to a celestial radiation source, under conditions often met in practice, is given by the two-dimensional convolution $A ** T$, where T is the temperature of the source as a function of direction and A is the response to a point source. A source is scanned in a raster pattern with a spacing of X milliradians between successive scans. Show that the data so obtained suffice to fill in the missing scans, provided $X < 500\ \lambda/w$, where λ is the wavelength and w is defined as follows. Make an orthogonal projection of the antenna aperture onto a plane perpendicular to the direction of the source; then w is the width of the projected aperture in the direction perpendicular to a raster line.

23. **Spoke diffraction pattern.** Consider a function $f(x,y)$ which consists of a line impulse $\delta(y)$ plus $N - 1$ copies rotated respectively $180° \, i/N$ where $i = 1, 2, \ldots, N - 1$. Thus there are N line impulses in all (or $2N$ "spokes"). Let $f(x,y) = \hat{f}(r,\theta)$. Write an expression for $\hat{f}(r,\theta)$. Find $\hat{F}(q,\theta)$, the two-dimensional Fourier transform of $f(x,y)$.

24. **Triangular aperture.** An equilateral triangular aperture many wavelengths in extent lies in a plane and has a uniform excitation of unity. Outside the triangle the excitation is zero. The angular spectrum is $P(l,m)$. Consider the contour $P(l,m) = 0.707P(0,0)$ (i.e., the half-power contour). (*a*) Does it remind you of a circle, a triangle, or some other shape? (*b*) Can you explain simply (other than by blind calculation) how it comes to possess the character that you have discovered? (*c*) Can you give a reason for thinking that there does or does not exist a null contour of $P(l,m)$ surrounding the main beam? ▷

25. **Apodization.** The aperture field distribution on a telescope of diameter $2a$ is characterized by $E_0(r) = \Pi(r/2a)$ and, associated with the discontinuous reduction to zero at $r = a$, there are strong diffraction rings, or optical sidelobes, that do not die away very rapidly. In fact, since

$$\Pi(r/2a) \text{ has Hankel transform } (2a)^2 \operatorname{jinc} 2aq,$$

where $\operatorname{jinc} q = (2q)^{-1}J_1(\pi q)$, and if $q \gg 1, J_1(q) \sim (2\pi q)^{-\frac{1}{2}} \cos(q - 3\pi/4)$, the sidelobe amplitude dies away as $q^{-3/2}$ and the power as q^{-3}. If we wish to make the sidelobes fall

to lower levels, we may remove the discontinuity and go even further, to remove the discontinuity in slope, and so on. Apertures treated by shading in this way are said to be apodized. For example, an aperture field distribution $(1 - r^2/a^2)\Pi(r/2a)$ is continuous but has discontinuous slope at $r = a$; $(1 - r^2/a^2)^2\Pi(r/2a)$ has continuous slope but discontinuous second derivative, and so on. In general, we can say that the kth derivative of $(1 - r^2/a^2)^k\Pi(r/2a)$ is discontinuous. As k increases, the aperture functions $(1 - r^2/a^2)^k\Pi(r/2a)$ thus generated exhibit progressively improved sidelobe reduction. They are called Sonine functions. Show that $(1 - r^2/a^2)^k\Pi(r/2a)$ has HT $\pi a^2 2^{k+1}k!(2\pi aq)^{-k-1}J_{k+1}(2\pi aq)$.

26. **Nulling interferometry.** A nearby star with an angular diameter of 0.001″ is to be inspected for the possible presence of a planet at an angular distance 0.5″ by use of a two-element interferometer. The star's radiation is 5000 times greater than the planet's. The two interferometer elements will be connected together in antiphase so that a null in the directional response can be pointed at the star. How many wavelengths should separate the interferometer elements in order that the presumed planet will lie in a direction of maximum reception of the interferometer? ▷

27. **Parabolic aperture distribution.** An aperture distribution is given by $E(\xi) = (1 - 4\xi^2/a^2)\Pi(\xi/a)$, where $\xi = x/\lambda$ and $a = 10$. (*a*) What is the overall width of the aperture in wavelengths? (*b*) Find the angular spectrum $P(s)$, normalized so that $P(0) = 1$. (*c*) Show that the first sidelobe occurs at $s = 0.01835$, or $\pm 1.05°$, from the beam axis and is 21 dB down relative to the main beam. ▷

16

Applications in Statistics

Fourier transform theory arises in statistics and in physical subjects involving random phenomena in several ways, many of them traceable to a certain convolution relation that plays a basic role. Application of the convolution theorem to this basic relation then introduces Fourier transforms and enables us to bring to bear the theory dealt with in this book.

The range of topics involved is extremely wide, embracing all those branches of physics and engineering in which random phenomena are considered. Our examples, however, can of necessity illustrate only a negligible fraction of the many possible applications. Suitable texts for related reading include Gardner (1986), Gray (1987), Leon-Garcia (1989), Papoulis (1965), and Parzen (1962).

We begin with a résumé of the terminology to be used. Examples of frequency distributions are the normal distribution (of errors), the Rayleigh distribution, the Poisson distribution, and other well-known functions describing the relative frequency with which some quantity assumes various possible values. In general we shall use the notation $P(x)$ for a frequency distribution; then the relative frequency, or probability, with which some quantity assumes values between x and $x + dx$ is $P(x)\,dx$, a dimensionless quantity. (In cases where dx has physical dimensions, $P(x)$ also has dimensions, which are the reciprocal of those of dx.)

Because of the variety of applications of statistical theory, we encounter many terms more or less synonymous with frequency distribution, such as probability distribution, probability law, probability density function, and distribution function.[1]

In the following section two problems are worked out which exhibit the convolution relation referred to above.

[1] The term "distribution function" is also used to denote cumulative frequency distribution, the integral of $P(x)$ from $-\infty$.

DISTRIBUTION OF A SUM

If we purchase a large number of 100-ohm resistors quoted as subject to a 10 percent tolerance, we shall find that the resistances range from 90 to 110 ohms, and that there are roughly as many in one 1-ohm interval as in any other. The variations are compatible with those to be expected if the resistors were drawn from an infinite supply containing equal numbers of resistors in any 1-ohm interval between 90 and 110 and none outside that range. This is a consequence of the method of manufacture and sorting by which the resistors are produced. (There is an insidious possibility that there will be no resistances between 99 and 101 ohms, those resistors having been sorted out and sold at a premium as precision resistors.) Let us suppose, therefore, that our stock of resistors has been drawn from a supply in which the frequency of occurrence of resistances between R and $R + dR$ is $P_1(R)\,dR$, where

$$P_1(R) = \frac{1}{20}\,\Pi\left(\frac{R-100}{20}\right).$$

In this formula $\Pi(R)$ is the unit rectangle function, and the factor $\frac{1}{20}$ allows for the requirement that

$$\int_0^\infty P_1(R)\,dR = 1.$$

The frequency distribution $P_1(R)$ is shown in Fig. 16.1.

If we also have a stock of 50-ohm resistors, also with a 10 percent tolerance, we have a second frequency distribution $P_2(R)$ given by

$$P_2(R) = \frac{1}{10}\,\Pi\left(\frac{R-50}{10}\right).$$

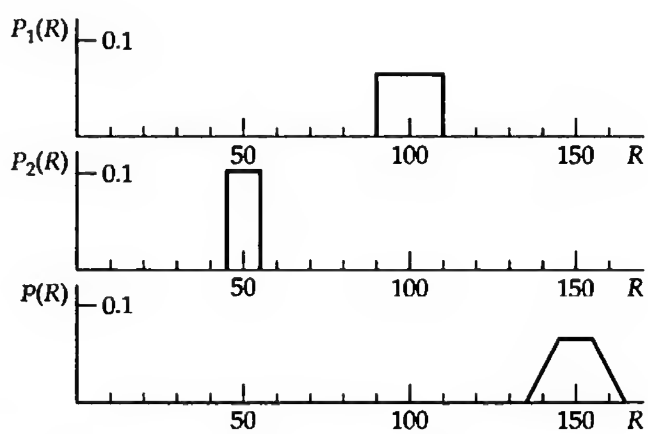


Fig. 16.1 **Frequency distribution of a resistance R formed by a series combination of two stock resistors, each having a tolerance of 10 percent.**

We propose to build an electronic circuit, in which the design calls for a 150-ohm resistance, by combining two of the stock resistors in series, and we require to know what tolerance should be ascribed to the composite element. In other words, we wish to know the distribution describing the frequency of occurrence of resistances lying between the minimum possible value of 135 ohms and the maximum of 165. Call this frequency distribution $P(R)$.

It is seen that both $P_1(R)$ and $P_2(R)$ enter into the determination of $P(R)$. The problem is solved as follows. Let the resistor picked from the 100-ohm stock have a resistance R'; then if the series combination is to have a total resistance R, the resistor from the 50-ohm stock will have to have resistance $R - R'$. The frequency of occurrence of a total resistance R will be the product of the frequencies of occurrence of R' in the first component and $R - R'$ in the second, integrated over the range of possibilities of R'; that is,

$$P(R) = \int_{-\infty}^{\infty} P_1(R')P_2(R - R')\, dR'.$$

In this integral the limits of integration could be put at 90 and 110, since that is the range of possible values of R' in the particular circumstances under consideration. Since, however, $P_1(R)$ has been defined to be zero outside the range of 90 to 110, the value of the integral is not affected by the use of infinite limits, and there is a gain in generality of the formula.

Inspection of the result now reveals that $P_1(R)$ and $P_2(R)$ enter into the constitution of $P(R)$ through the standard convolution integral; in fact, in asterisk notation,

$$P = P_1 * P_2.$$

This is an instance of the basic convolution relation between the frequency distribution describing the sum of two quantities and the frequency distributions of the given quantities. In this particular case the convolution can readily be performed, and it is found that the composite resistance has a mean value of 150 ohms and is distributed trapezoidally as shown in Fig. 16.1.

This trapezoidal response is, of course, what would result from scanning a uniform stripe 20 units wide on a sound track with a slit 10 units wide; as the slit moved onto the stripe, there would be a linear change in response, then an interval of unchanged response while the slit was entirely contained within the limits of the stripe, and then a linear return to the initial state. The duration of each of the three phases would be 10 units. A similar analogy exists with any other situation involving convolution. For example, a rectangular pulse lasting 20 units of time, if put into a filter whose impulse response was a rectangular pulse lasting 10 units of time, would produce the trapezoidal output illustrated.

Among the properties of convolution is the additive property of the abscissas of the centroids, and it is seen that the mean value of the composite resistance is indeed the sum of the means of the components. Also, the area under the convolution is the product of the areas under the convolved functions; since both $P_1(R)$ and $P_2(R)$ had unit area to begin with, the same should be true of $P(R)$, and inspection confirms that this is so. Variances are also additive under convolution;

from this it can be seen that the standard deviation of the composite resistance from its mean represents a smaller percentage than for the components. In this sense the tolerance has been improved but, of course, the spread between extremes is still 10 percent. If, however, the original distributions had been more rounded, or if more elements were connected in series, then the tendency toward a Gaussian result in accordance with the central-limit theorem would be more advanced, and the use of the standard deviation instead of the extreme range as a measure of spread would be quite appropriate.

As a second example, consider a barrel full of money, half consisting of dollar bills, the rest being fives and tens and a few twenties. To be more precise, let the frequency of occurrence of n-dollar bills be proportional to the nth term in the sequence

$$\{10\ 0\ 0\ 0\ 6\ 0\ 0\ 0\ 0\ 3\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 1\}.$$

Since the sum of the terms of the sequence is 20, it will be convenient to deal instead with a normalized sequence whose sum is unity:

$$\{\tfrac{10}{20}\ 0\ 0\ 0\ \tfrac{6}{20}\ 0\ 0\ 0\ 0\ \tfrac{3}{20}\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ \tfrac{1}{20}\}.$$

Let the nth term of the sequence be denoted by $p_1(n)$; it is the frequency of occurrence of n-dollar bills in the barrel.

If a single bill is drawn from the barrel, this sequence describes the relative frequencies with which bills of different denominations may be expected. Now if two bills are drawn, and the amounts added together, we may ask what will be the corresponding sequence describing the distribution of the total sum drawn. Let $p(n)$ be the nth term of the wanted sequence. If the first draw is m dollars, then the second must be $n - m$ if the total sum drawn is to be n dollars. By the product rule for the joint occurrence of independent events, $p_1(m)p_1(n - m,)$ is the frequency with which we may expect to draw just m dollars on the first draw and $n - m$ on the second. Summing over all possible first choices m, we have

$$p(n) = \sum_m p_1(m)p_1(n - m).$$

Thus the wanted sequence is the serial product of the given sequence with itself, or, in asterisk notation,

$$\{p\} = \{p_1\} * \{p_1\}.$$

Figure 16.2 shows the calculation being carried out numerically by writing the given sequence upward on a movable strip of paper. The strip is shown in position for calculating the value of $\frac{120}{400}$ opposite the arrow; this is the frequency with which we might expect to draw six dollars. The values in parentheses still remain to be calculated at this point. The full result is

$$\{0\ \tfrac{100}{400}\ 0\ 0\ 0\ \tfrac{120}{400}\ 0\ 0\ 0\ \tfrac{36}{400}\ \tfrac{60}{400}\ 0\ 0\ 0\ \tfrac{36}{400}\ 0\ 0\ 0\ 0\ \tfrac{9}{400}\ \tfrac{20}{400}$$
$$0\ 0\ 0\ \tfrac{12}{400}\ 0\ 0\ 0\ 0\ \tfrac{6}{400}\ \cdots\ \tfrac{1}{400}\}.$$

As a numerical check it may be verified that the sum of the terms is unity.

3
0
0
0
0
6
0
0
0
10

n	$p_1(n)$	$p(n)$
	+20	+400
1	10	0
2	0	100
3	0	0
4	0	0
5	6	0
6	0	120
7	0	0
8	0	0
9	0	0
10	3	(36)
11	0	(60)
12	0	0
13	0	0
14	0	0
15	0	(36)
16	0	0
17	0	0
18	0	0
19	0	0
20	1	(9)
21		(20)
22		0
23		0
24		0
25		(12)
26		0
27		0
28		0
29		0
30		(6)
31		0
32		0
33		0
34		0
35		0
36		0
37		0
38		0
39		0
40		(1)

Fig. 16.2 Calculating the serial product of $p_1(n)$ with itself.

The very close similarity between this example and the earlier one dealing with resistors will be apparent. In one case a wanted frequency distribution depends on two given distributions through convolution, and in the other through a serial product. However, if in the resistor problem numerical computation were resorted to in a case where the given distributions consisted of data not amenable to integration by analytic methods, we would compute a serial product based on sufficiently close-spaced data. Thus the numerical approach to both problems is the same.

Conversely, there is nothing gained by making a distinction between the two problems in algebraic treatment; in this case, however, we treat both in terms of

convolution. Thus the contents of the barrel may be described by a frequency distribution over a continuous variable x by writing

$$P_1(x) = \tfrac{10}{20}\delta(x-1) + \tfrac{6}{20}\delta(x-5) + \tfrac{3}{20}\delta(x-10) + \tfrac{1}{20}\delta(x-20).$$

Figure 16.3 illustrates this by means of the convention that an impulse symbol is represented by an arrow of height equal to the strength of the impulse. The figure also shows the distribution for the sum of two draws, namely $P(x)$, which may be written

$$\begin{aligned} P(x) &= P_1(x) * P_1(x) \\ &= \tfrac{1}{400}[100\delta(x-2) + 120\delta(x-6) + 36\delta(x-10) + 60\delta(x-11) \\ &\quad + 36\delta(x-15) + 9\delta(x-20) + 20\delta(x-21) \\ &\quad + 12\delta(x-25) + 6\delta(x-30) + \delta(x-40)]. \end{aligned}$$

It may be verified that the integral of both $P_1(x)$ and $P(x)$ is unity.

We have now seen the basic convolution relation, which may be restated briefly as follows. If two quantities can assume values of x describable by frequency distributions $P_1(x)$ and $P_2(x)$, respectively, then their sum is described by a frequency distribution $P(x)$ given by

$$P(x) = P_1(x) * P_2(x).$$

We have assumed that neither of the two quantities is influenced by the choice of the other. But if the barrel of money had been filled by throwing in great wads of each denomination and the contents had not been thoroughly stirred, then the results could be different; again, if there were only one twenty-dollar bill in the barrel, then drawing it would radically influence the second draw. Before the convolution relation is applied, it is always necessary to consider whether the two draws are independent, or whether the product rule is applicable.

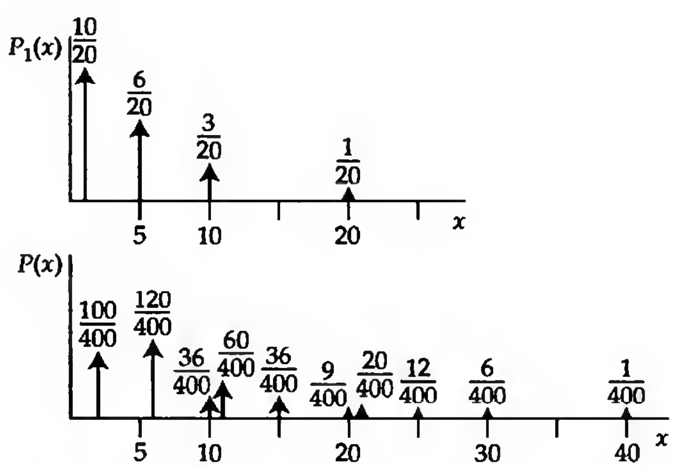


Fig. 16.3 **Frequency distributions expressed in terms of impulse symbols.**

CONSEQUENCES OF THE CONVOLUTION RELATION

Let $P(x)$ be the convolution of $P_1(x)$ and $P_2(x)$; that is,

$$P(x) = \int_{-\infty}^{\infty} P_1(x-u)P_2(u)\,du,$$

and let m, m_1, and m_2 be the abscissas of the centroids of the three distributions P, P_1, and P_2; that is,

$$m = \frac{\int_{-\infty}^{\infty} xP(x)\,dx}{\int_{-\infty}^{\infty} P(x)\,dx}, \ldots .$$

Since it has been shown that the abscissas of the centroids are additive under convolution, it follows that

$$m = m_1 + m_2.$$

If P, P_1, and P_2 are frequency distributions, then m, m_1, and m_2 are the mean values for the three distributions. Therefore the theorem quoted above corresponds, in statistics, to the result that the mean of the sum of two random variables is the sum of the two respective means.

According to a second theorem, variances are also additive under convolution. Let

$$\sigma^2 = \frac{\int_{-\infty}^{\infty} (x-m)^2P(x)\,dx}{\int_{-\infty}^{\infty} P(x)\,dx}, \ldots .$$

Then

$$\sigma^2 = \sigma_1^2 + \sigma_2^2.$$

The interpretation of this result in terms of statistics is well known.

If three random quantities are added together, and their respective frequency distributions are $P_1(x)$, $P_2(x)$, and $P_3(x)$, then the frequency distribution of the sum of the first two is $P_1 * P_2$, and the frequency distribution of the sum of all three must be $(P_1 * P_2) * P_3$. But by the associative property of convolution it is not necessary to retain the parentheses, and therefore the distribution of the sum of the three quantities $P(x)$ is given by

$$P = P_1 * P_2 * P_3.$$

It also follows that

$$m = m_1 + m_2 + m_3$$

and

$$\sigma^2 = \sigma_1^2 + \sigma_2^2 + \sigma_3^2.$$

This result clearly extends to the sum of any number of random quantities.

From the central-limit theorem we know that if several functions are convolved together, the result approaches a Gaussian distribution, provided that certain conditions are fulfilled. Therefore it follows that if several random quantities are added, then the frequency distribution of the sum will approach a Gaussian distribution. The mean of the distribution will be the sum of the component means, and the variance will be the sum of the component variances. If sufficient quantities are involved to ensure that the Gaussian result is approached within an accuracy that is sufficient for some purpose, then the additive properties of the mean and variance suffice to fix the parameters of the Gaussian distribution.

A final consequence of the convolution relation is that known properties of Fourier transforms may be brought to bear on statistics through the convolution theorem. This will now be considered.

THE CHARACTERISTIC FUNCTION

The characteristic function $\phi(t)$ associated with a frequency distribution $P(x)$ is defined by the relation

$$\phi(t) = \int_{-\infty}^{\infty} e^{itx} P(x)\, dx$$

and is thus the plus-i Fourier transform of the frequency distribution.

By transforming the relation

$$P = P_1 * P_2$$

we find

$$\phi(t) = \phi_1(t)\phi_2(t).$$

In other words, by the convolution theorem, simple multiplication of the characteristic functions gives the characteristic function of the frequency distribution of a sum.

Characteristic functions possess the following properties:

a. $\phi(0) = 1$ [since $\int_{-\infty}^{\infty} P(x)\, dx = 1$];

b. $\phi(t)$ is hermitian [since $P(x)$ is real]; that is, the real part of $\phi(t)$ is even, and the imaginary part odd;

c. $|\phi(t)| \leqslant 1$.

In addition, the real and imaginary parts of the characteristic function may not be independent. If, for example, $P(x)$ is zero for $x < 0$, the real and imaginary parts of $\phi(t)$ will form a Hilbert transform pair. If $P(x)$ is zero over some other semi-infinite range then an appropriate application of the shift theorem will produce a similar interconnection.

THE TRUNCATED EXPONENTIAL DISTRIBUTION

As an illustration of the convolution relation, and to lay a basis for later considerations of noise waveforms, we consider the following frequency distribution:

$$P(x) = X^{-1}e^{-x/X}H(x),$$

where $H(x)$ is the Heaviside unit step function (Fig. 16.4). This will be referred to as the truncated exponential distribution. The mean m and variance σ^2 are given, respectively, by

$$m = X$$

$$\sigma^2 = X^2.$$

This frequency distribution arises in a fundamental way. Consider a completely random event, the best example of which is the spontaneous radioactive disintegration of an atomic nucleus. The occurrence of telephone calls, or the emission of electrons from a cathode might furnish other examples under certain conditions. Figure 16.5 shows a record of such individual events. It has been shown by observation that an unstable atom is as likely to disintegrate at any one moment as it is at any other moment; thus the number of disintegrations occurring per second in a radioactive mass is proportional, on the average, to the number of unstable atoms present. We shall assume that this number is so large that it is not noticeably depleted by the occurrence of the disintegrations.

Examination of Fig. 16.5 shows that the interval between one event and the next has a certain average value corresponding to the average rate of occurrence, but also shows that intervals shorter than average are more frequent than longer ones. In fact, the shorter the interval, the more frequent, right down to zero interval. This fact is connected with a folk saying that misfortunes occur in pairs.

We now show that the frequency of occurrence of an interval between x and $x + dx$ is $P(x)\,dx$, where

$$P(x) = X^{-1}e^{-x/X}H(x),$$

and X is the mean interval between events.

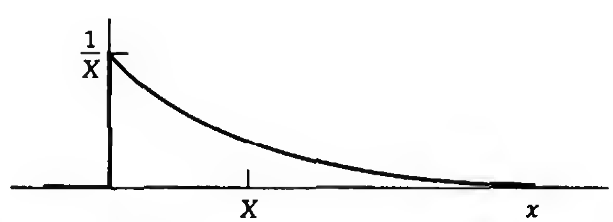


Fig. 16.4 **The truncated exponential distribution with mean *X*.**

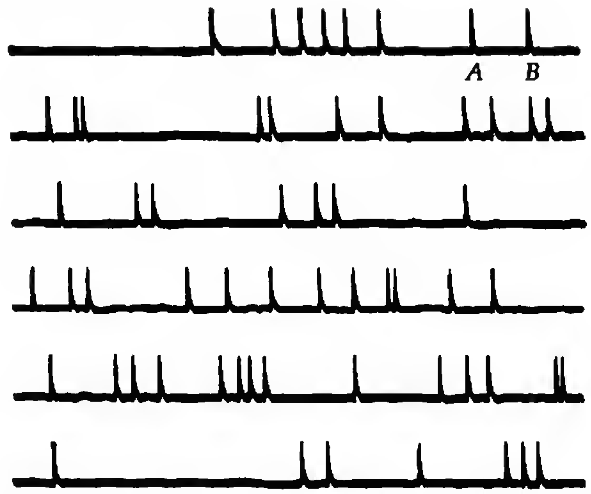


Fig. 16.5 **A record of events occurring at random but at a constant average rate. The average interval is *AB*.**

Because the average interval between events is X, the average rate of occurrence is X^{-1} per unit of x. Therefore the frequency of occurrence in a brief space Δx will be $X^{-1}\,\Delta x$; and the frequency of nonoccurrence will be $1 - X^{-1}\,\Delta x$. Suppose that, following one event, an interval x elapses before the occurrence of the next. Let x be N times Δx. The frequency with which the random event will fail to occur in each of N successive spaces Δx, to be followed by an occurrence in the succeeding space Δx, is by definition approximately $P(x)\,\Delta x$, and it is given by

$$P(x)\,\Delta x = (1 - X^{-1}\,\Delta x)^N X^{-1}\,\Delta x \qquad x > 0.$$

In the limit as $\Delta x \to 0$,

$$\begin{aligned} P(x) &= \lim_{\Delta x \to 0} (1 - X^{-1}\,\Delta x)^{x/\Delta x} X^{-1} \\ &= X^{-1} e^{-x/X} \qquad x > 0. \end{aligned}$$

The frequency of an event with respect to its predecessor declines exponentially with delay (which cannot be negative). Hence

$$P(x) = X^{-1} e^{-x/X} H(x).$$

The truncated exponential distribution arises wherever events occur absolutely randomly in time or in one-dimensional space, and then it describes the frequency distribution of the interval between one event and the next.

It also arises commonly in another connection. The events shown in Fig. 16.5 were identical events, but it often happens that events not only occur at random but have various sizes. Size distributions resembling the exponential are common, for example, among geophysical phenomena where small earthquakes, small showers of rain, and the like tend to be frequent and the larger events rare. Such

size distributions often occur because the phenomena are triggered at random and then release a quantity of something (such as elastic energy, or water) that has been accumulating steadily since the preceding event. If the size of the event is proportional to the elapsed interval, and the event is as likely to occur at one moment as at any other, then the frequency distribution of sizes will follow the truncated exponential law exactly.

Other quite distinct ways exist that give rise to the same distribution. One important case comes up later in connection with the detection of noise waveforms.

THE POISSON DISTRIBUTION

This discussion takes as its point of departure the truncated exponential distribution, denoted by $E(x)$, where

$$E(x) = e^{-x}H(x).$$

By applying the convolution relation to describe the frequency distribution of a sum of quantities taken successively from distributions $E(x)$, we obtain an understanding of the Poisson distribution. This is important in itself, but in addition it furnishes excellent examples of standard procedures with frequency distributions. In particular, we gain a detailed view of a transition into a Gaussian distribution as envisaged by the central-limit theorem.

To lend an air of reality to the development, we shall express it in terms of a physical situation. Thus let us assume that in a certain locality the frequency $P(x)\,dx$ with which one day's rain is between x and $x + dx$, is given by

$$P(x) = E(x).$$

The average daily rainfall is thus one unit.

How much rain falls on two days? Since it is the sum of two quantities drawn from a known frequency distribution, use of the convolution relation is indicated. First, it is necessary to verify that the rainfall on any day is independent of that on the previous day. In many localities this is not so, each day tending to be like the previous day; but we are going to deal here with an unpredictable climate. Then the frequency distribution of two-day totals is

$$P * P.$$

Performing the indicated convolution, we find

$$\begin{aligned} P * P &= E(x) * E(x) \\ &= \int_{-\infty}^{\infty} e^{-x'}H(x')e^{-(x-x')}H(x - x')\,dx' \\ &= \int_0^x e^{-x'}e^{-(x-x')}\,dx' \\ &= xE(x). \end{aligned}$$

The frequency distribution $E * E$ of two-day totals is thus like the impulse response of a critically damped instrument (see Fig. 16.6 for E and $E * E$). This is different from the truncated exponential in that the most frequent two-day total is no longer zero; in fact there is a marked alteration in character in the direction of a normal or Gaussian distribution. The mean two-day total is two units; the mode is at one unit.

The distribution of n-day totals is the n-fold self-convolution of $E(x)$,

$$[E(x)]^{*n}.$$

We can easily evaluate this distribution for $n = 3$ by using the result for $n = 2$, obtaining

$$[E(x)]^{*3} = \frac{x^2}{2!}e^{-x}H(x).$$

This suggests that the n-day totals have a distribution

$$\frac{x^{n-1}}{(n-1)!}E(x);$$

if this is so, then the $(n + 1)$-day total would be

$$[E(x)] * \left[\frac{x^{n-1}}{(n-1)!}E(x)\right] = \int_0^x \left[\frac{u^{n-1}}{(n-1)!}e^{-u}\right]e^{-(x-u)}\,du$$
$$= \frac{x^n}{n!}E(x).$$

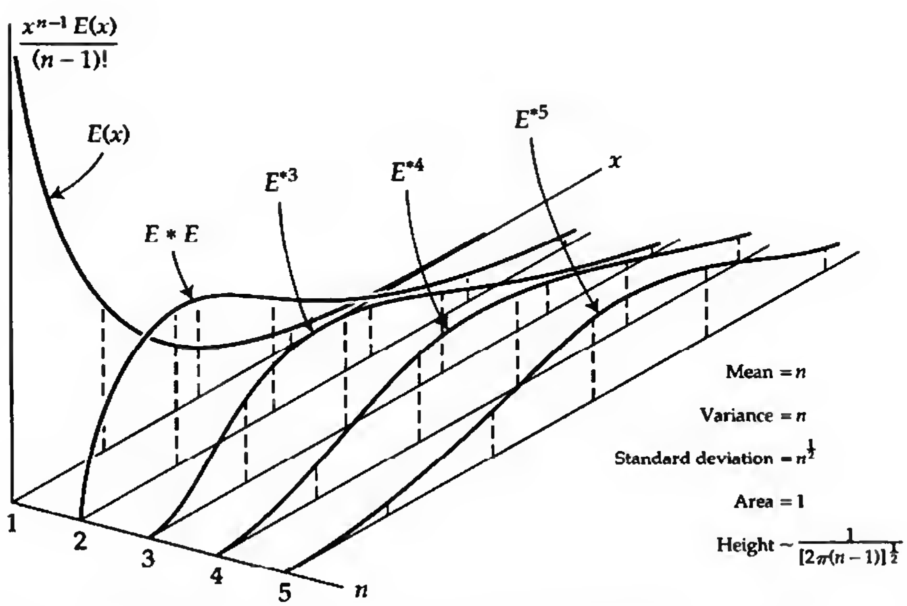


Fig. 16.6 The Pearson Type III distribution exhibited as successive self-convolutions of the truncated exponential $E(x)$.

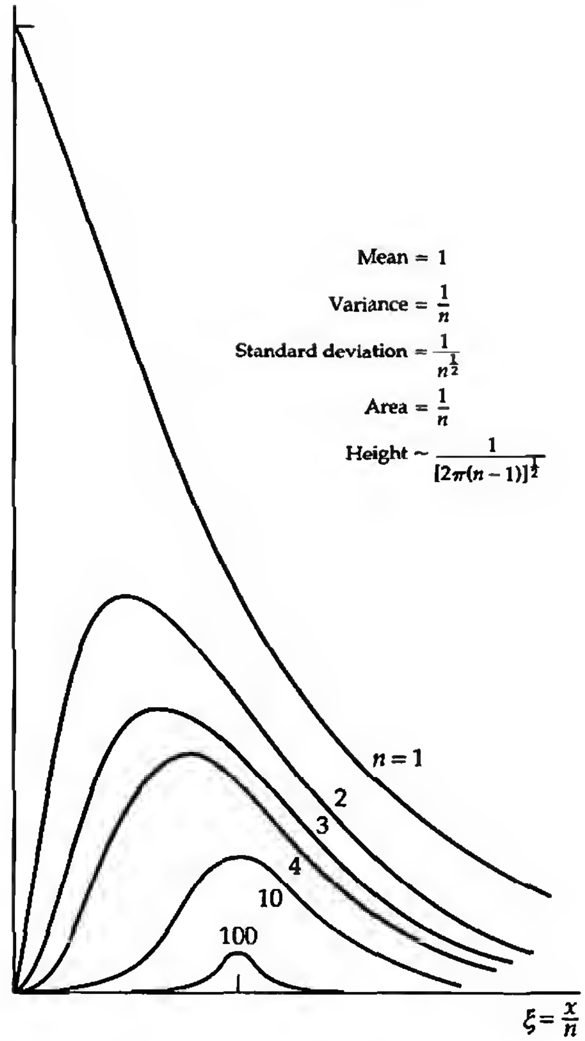


Fig. 16.7 E^{*n} **normalized to** $\langle x \rangle = 1$ **to show the approach to Gaussian form.**

Thus if the conjecture is correct for n-day totals, it is also correct for $(n + 1)$-day totals, and so on. But it is known to be correct for $n = 3$; hence it is correct for $n = 4$, and so on *ad infinitum*.

The distribution

$$\frac{x^n}{n!} E(x) = [E(x)]^{*n+1}$$

shown in Fig. 16.6 may be referred to as a Pearson Type III distribution. It represents the frequency with which the sum of $n + 1$ exponentially distributed quantities with unit mean value has a value x.

By the central-limit theorem we know that for large values of n we shall have approximately Gaussian distributions.

The mean value of x is equal to $n + 1$, and the variance is equal to $n + 1$, as a consequence of the theorems that variance and the abscissa of the center of gravity are additive under convolution. (The mode can be shown to fall at $x = n$ by differentiation.)

For large n, the standard deviation about the mean is $n^{-\frac{1}{2}}$ times the mean. Thus if we express the distribution in terms of $\xi = x/n$, so as to refer all the distributions to the same mean, we have

$$\frac{x^{n-1}}{(n-1)!}E(x) = \frac{n^{n-1}}{(n-1)!}\xi^{n-1}e^{-n\xi}H(\xi).$$

We expect this formula to exhibit the contraction to an ever-narrower Gaussian form, as in Fig. 16.7; that is, we expect that

$$\frac{n^{n-1}}{(n-1)!}\xi^{n-1}e^{-n\xi}H(\xi) \rightarrow \frac{1}{(2\pi n)^{\frac{1}{2}}}e^{-\frac{1}{2}n(\xi-1)^2}.$$

Putting $\epsilon = \xi - 1$, and making use of the Stirling formula for factorials of large numbers

$$n! = (2\pi n)^{\frac{1}{2}}n^n e^{-n},$$

x \ n	0	1	2	3	4	5
0	1.000					
1	.368	.368	.184	.061	.015	.003
2	.135	.271	.271	.180	.090	.036
3	.050	.149	.224	.224	.168	.101
4	.018	.073	.146	.195	.195	.156
5	.007	.034	.084	.140	.175	.175

Mean = x

Variance = x

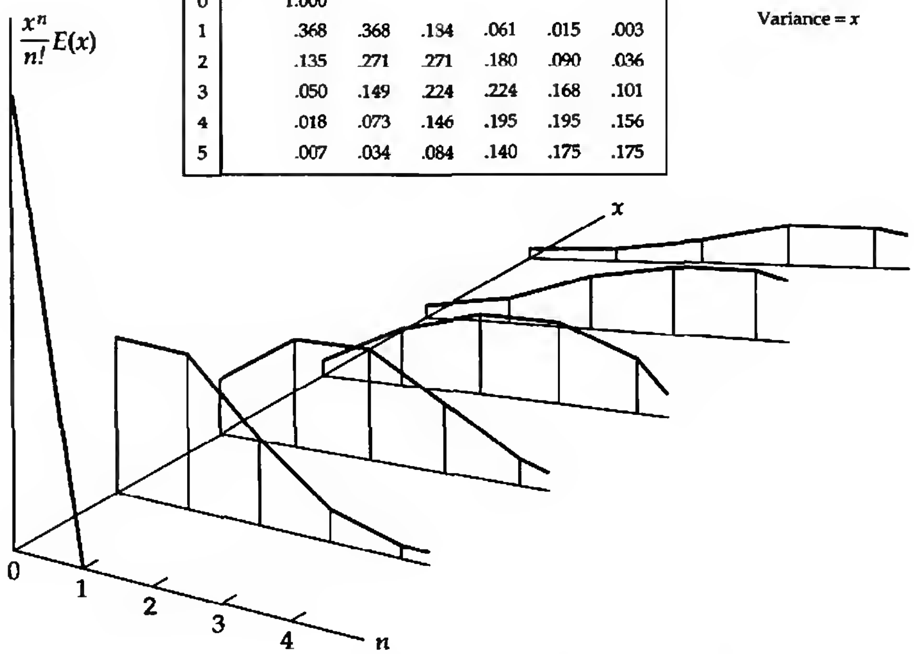


Fig. 16.8 **The Poisson distribution regarded as a continuous set of histograms.**

the left-hand side reduces to

$$\frac{n^n}{n!}(1+\varepsilon)^{n-1}e^{-n(1+\varepsilon)} = \frac{1}{(2\pi n)^{\frac{1}{2}}}(1+\varepsilon)^{n-1}e^{-n\varepsilon},$$

which, for small values of ϵ, expands to

$$\frac{1}{(2\pi n)^{\frac{1}{2}}}[1 - \epsilon - \tfrac{1}{2}(n-2)\epsilon^2 + \ldots]$$

and thus matches, as far as these first terms go, a Gaussian distribution centered on $\xi = 1 - 1/n$.

If events occur at random as in Fig. 16.5 with an average interval of unity, the probability that precisely n events will occur in the interval x depends first on $[E(x')]^{*n}\,dx'$ (the probability that the nth event from now will occur in the interval $x' \pm \frac{1}{2}dx'$, $x' < x$) and then on the probability that the one following those will occur *after* a further interval $x - x'$. This second probability is the integral of $E(x)$ from $x - x'$ to infinity and is thus equal to $E(x - x')$. Multiplying these two probabilities and integrating over the range of possibilities of x', we have

$$\int_0^x E(x - x')[E(x')]^{*n}\,dx' = [E(x)]^{*(n+1)} = \frac{x^n}{n!}E(x).$$

This is precisely the function we have been studying as a probability density per unit of x; as a discrete probability depending on n, however, we know it as the Poisson distribution (Fig. 16.8).

BIBLIOGRAPHY

Gardner, W. A.: "Statistical Spectral Analysis," Prentice-Hall, Englewood Cliffs, New Jersey, 1988.

Gray, R. M.: "Probability, Random Processes, and Ergodic Properties," Springer, New York, 1987.

Leon-Garcia, A.: "Probability and Random Processes for Electrical Engineering," Addison-Wesley, Reading, Massachusetts, 1989.

Papoulis, A.: "Systems and Transforms with Applications in Optics," McGraw-Hill, New York, 1968.

Parzen, E.: "Modern Probability Theory and Its Applications," Wiley, New York, 1960.

PROBLEMS

1. **Sum of reciprocals.** The distribution of resistance in a batch of nominally identical resistors is $P(x)$. Calculate the distribution of resistance of parallel combinations of two resistors from the batch.

2. Show that the frequency distribution of the complex impedance $Z = X + iY$ obtained by connecting in series a resistance R and inductance L, each equally likely to be anywhere within one percent of the nominal value, is given by

$$P(X,Y) = \left[50\Pi\left(\frac{X - R}{0.02R}\right)\delta(Y)/R\right] ** \left[50\Pi\left(\frac{Y - \omega L}{0.02\omega L}\right)\delta(X)/\omega L\right],$$

where the double asterisk represents two-dimensional convolution, and sketch this distribution on the complex plane of Z.

3. **Beating the odds.** It is proposed to set up an absolute standard 1-megohm resistor of unusual precision, for delicate measurements on insulators, by taking 10,000 ordinary 1 percent 100-ohm resistors and connecting them in series by means of mercury cups of negligible resistance. The argument given is as follows. If n resistors from a batch of mean resistance R and variance σ^2 are connected in series, the combined resistance will be, on the average over many trials, nR, and the variance will be $n\sigma^2$. The standard deviation about the mean will be $n^{\frac{1}{2}}\sigma$. Consequently, if σ/R was originally of the order of 10^{-2}, it will be of the order of

$$\frac{n^{\frac{1}{2}}\sigma}{nR} = \frac{n^{-\frac{1}{2}}\sigma}{R} = (10{,}000)^{-\frac{1}{2}}\frac{\sigma}{R} = 10^{-2} \times 10^{-2} = 10^{-4}$$

for 10,000 resistors in series. Explain the fallacy in thinking that a resistance equal to 1 megohm within one part in 10^4 could be achieved in this way. ▷

4. **Bernoulli distribution.** The probability of emission of n electrons in 1 microsecond from a microscopic area of a cathode is

$$0.9\delta(n) + 0.1\delta(n - 1).$$

Calculate the probability of emission of n electrons in one second from a cathode that is more extensive by a large factor M.

5. **Poisson distribution.** If a random event occurs 0.1 times per microsecond on the average, what is the frequency of occurrence of just two events in 1 microsecond?

6. **Distribution of a sum.** Calculate the frequency distributions of the totals thrown with (*a*) a pair of dice, (*b*) three dice.

7. **Distribution of a sum.** A flip-flop circuit has two stable states, A and B. It changes from one state to the other whenever it is pulsed. If the pulses arrive at an average rate of 10 per second, but at completely random times (as in Fig. 16.5), calculate the frequency distribution for the time interval between successive transitions from state A to state B. What is the most frequent value?

8. **Characteristic function.** Show that if $\varphi(t)$ is a characteristic function, then it must be expressible as the autocorrelation function of some other function. Show that $2t$ cosech πt can be a characteristic function, and determine two or more functions of which it is the autocorrelation function. ▷

9. **Spatial probability.** A target whose location is better known in latitude than in longitude is fired upon by a weapon which is more accurate in azimuth than in range.

Calculate the two-dimensional probability distribution of the miss distance. Is it more favorable (for the one firing) to fire from the north or from the east?

10. **Moments.** It is known that the mean and variance are additive under conditions where the probability distribution of a sum is given by a convolution relation. Is this true of the third moment? ▷

11. **Meteor-trail communication.** A point-to-point communication system utilizes signal bursts due to specular reflection from the transient ionized trails left by meteors. If the transmitter is left on, the signal amplitude at the receiver consists of bursts of the form $A = A_0 e^{-(t-t_i)/\tau} H(t - t_i)$ occurring at random epochs t_i. The number of bursts with amplitude greater than A_1 is given by $N = cA_1^{-a}$ during a certain period when c and a are constant. If communication is possible when the signal exceeds the threshold A_1, show that the channel is open for a time T given by

$$T = \frac{\tau}{a} c A_1^{-a}$$

and that the fraction of usable bursts with durations exceeding t_1 is $e^{-at_1/\tau}$.

12. A Bernoulli trial is one whose outcome x is described by the probability distribution $q\,\delta(x) + p\,\delta(x-1)$, where $q = 1 - p$. In terms of discrete probabilities this may be written $\{q \;\; p\}$. The sum of the outcomes after n independent trials has a distribution $\{q \;\; p\}^{*n}$, called the binomial distribution [because its terms $\binom{n}{x} q^{n-x} p^x$ are the same as in the expansion of the nth power of the binomial expression $q + p$]. Show that the characteristic function of the binomial distribution is $[q + p \exp it]^n$ and deduce that if two independent quantities have binomial distributions then their sum has a binomial distribution provided that p is the same for each.

13. Show that the characteristic function of the Poisson distribution $(x^n/n!)e^{-x}$ is $\exp\{x[\exp it - 1]\}$ and deduce that if two independent quantities have Poisson distributions with means x_1 and x_2, then their sum has a Poisson distribution with mean $x_1 + x_2$. ▷

14. Verify that the terms of the Poisson distribution $(x^n/n!)\exp(-x)$, when summed for $n = 0, 1, 2 \ldots$, add up to unity. Show that, when x is a positive integer, there are always two equal terms.

15. **Probability distribution of sum.** The chances of hitching a ride from the campus to the freeway on Sunday night are such that the probability is $P_1(t)\,dt$ that it will take a time between t and $t + dt$ to get a lift, counting from the time the thumb is extended. Reports indicate that $P_1(t)$ is approximated by $P_1(t) = 0.025 \exp(-t/40)$, where t is in minutes. The travel time to the freeway averages 10 minutes but itself varies over a narrow range with a standard deviation of 2 minutes. Waiting time at the freeway is described by $P_3(t) = 0.1 \exp(-t/10)$ and the travel time to the airport is 25 minutes ± 5 (standard deviation). What is the mean travel time from the campus to the airport by this mode of transportation and what is the standard deviation?

16. **Distribution of reciprocal.** The frequency of occurrence of resistances between R and $R + dR$ is $0.05\Pi(0.05R - 5)$, but when they are tested on an instrument which measures conductance it is found that the distribution does not appear to be flat-topped. What is the probable explanation? ▷

17. **Bernoulli's theorem, central limit theorem.** The impulse pair $p\delta(x) + q\delta(x - X)$, where p and q are both positive and $p + q = 1$, may be convolved with itself several times to produce a string of impulses $[p\delta(x) + q\delta(x - X)]^{*n}$ whose strengths, as $n \to \infty$, delineate a certain Gaussian function. Show that the multiple self-convolution with n factors approaches

$$(2\pi npq)^{\frac{1}{2}} e^{-(x - nqX)^2/2npqX^2}\, \text{III}(x/X).$$

18. **Central limit theorem, Bernoulli's theorem.** The quantities p and q in Bernoulli's theorem are necessarily positive because they represent probabilities. But if p and q were of opposite sign, what could be said about the n-fold self-convolution of $p\delta(x) + q\delta(x - X)$?

17

Random Waveforms and Noise

The whole of physics is permeated by naturally occurring random phenomena, and we may draw some examples from an account of a straightforward physical observation. Figure 17.1 shows a record obtained by allowing the extragalactic radio source Cygnus *A* to pass through the beam of a radiotelescope. Clearly this radio source consists of at least two parts, but it is important to know whether there is also a halo associated with the two obvious components. The existence of a third source, broader and weaker than the first two, cannot, however, be revealed by studying this record, because of the presence of unwanted wiggles in the record.

This is a typical example of detection of a signal against a noise background. The sources of the fluctuations in this example are well understood; microwave radiation emitted thermally from the earth and from the walls of the antenna transmission line enters the receiver, and the receiver itself generates radiation; the background emission of the galaxy contributes a tiny amount, but the source under study itself contributes significantly even when it is out of the main beam of the antenna. All this unwanted radiation adds up, and the total is not constant. The *fluctuations*, as the illustration shows, are smaller than the wanted signal, although the *total* unwanted power is actually very much greater. Such fluctuations often reveal characteristic behavior that is quite independent of their precise physical cause, and this chapter will be concerned with these universal phenomena. However, it should certainly be recorded here that actual observations also often reveal behavior that is inexplicable from the standpoint of prior theory and can only be interpreted after observational experience; these peculiar phenomena, which may become the dominant ones as the more understandable fluctuations are gradually reduced by appropriate measures, are too diverse to be studied from the literature.

In our example, the wanted radiation itself also reveals random aspects; in fact the signal that we are extracting from the noise background is itself random noise. It is almost completely randomly polarized, which means that the electric

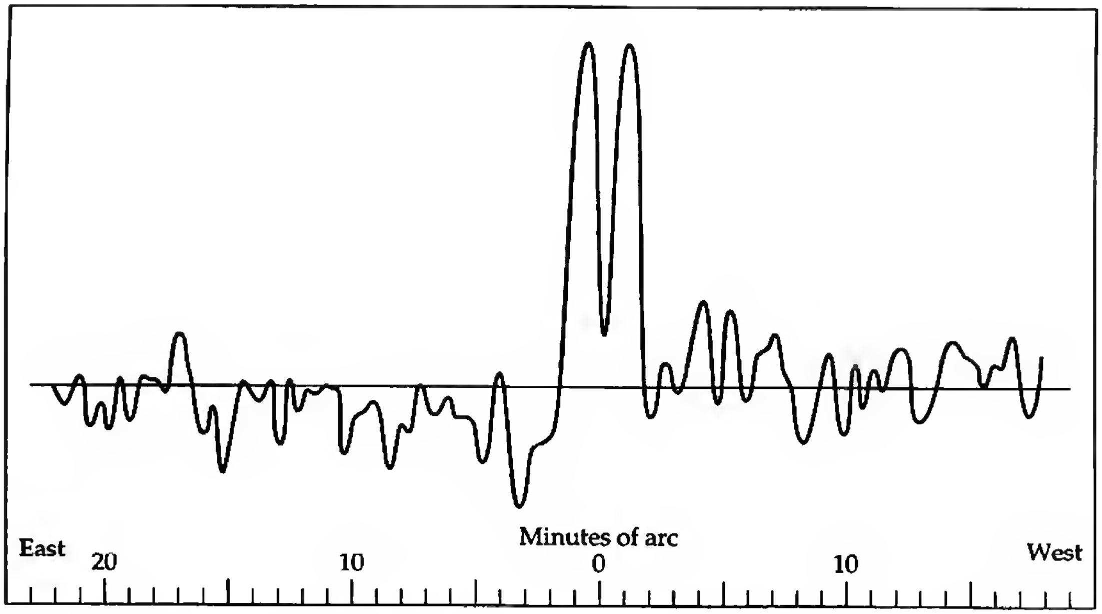


Fig. 17.1 **Record of power received from the extragalactic radio source Cygnus *A* *passing through the beam of a radio telescope, the first instrument to* generate a radio beam (52 arcseconds, one dimensional) as good as that of the human eye (Swarup, Thompson and Bracewell, 1963).**

field in the wavefront is about as likely to point in one direction as any other. Furthermore, the amplitude of the electric field of the signal fluctuates in a characteristic way as time elapses. The fluctuations of the horizontal and vertical components of the field are identical in character but almost completely different in detail.

For reading related to this chapter see Gardner (1988), Goodman (1996), Papoulis (1968), and Ripley (1981).

DISCRETE REPRESENTATION BY RANDOM DIGITS

To study the characteristics of noise in the various forms discussed above, we first construct a particular noise waveform by a purely numerical procedure and become familiar with its properties. In this example we work with sequences of numbers which may be deemed to represent successive samples of a band-limited waveform; therefore we need not hesitate at any time to graph the sequence and draw a smooth curve through the points. To draw such a curve we may interpolate in accordance with the rules given in connection with the sampling theorem in Chapter 10.

Suppose that digits from 0 to 9 can be drawn from a supply, any one digit being as likely to be drawn as any other. For instance, we could take the random

numbers distributed uniformly between 0 and 1 that are provided by a computer language, multiply by 10, and take the integer part. But for the sake of a concrete example, we shall use here the successive digits in the decimal representation of π. We know that

$$\pi = 3.14159265358979323846264338327950288419716939937510\ldots.$$

These first 32 digits are graphed in Fig. 17.2. It has been shown that in almost all decimals the various digits occur with uniform frequency; that is, each occurs in about one-tenth of the possible places; and in the case of π this question has been the subject of investigation out to extreme numbers of decimals. Certainly in the finite run illustrated the number of occurrences is not the same for each digit (the histogram on the left gives the totals), but there is no way of proving that this finite sequence is incompatible with having been drawn from a uniform supply.

Let y_t be the value of the tth digit; that is, $y_1 = 3, y_2 = 1, y_3 = 4, \ldots$. We may also refer to the sequence $\{y_t\}$ and write

$$\{y_t\} = \{3\,1\,4\,1\,5\,9\ldots\}.$$

The quantity t may be regarded as time, and it may be supposed that digits are drawn at instants regularly spaced at one unit of time. The mean m of the first N digits is given by

$$m = \frac{y_1 + y_2 + y_3 + \ldots + y_N}{N} = \frac{1}{N}\sum_{t=1}^{N} y_t;$$

for instance, when $N = 32$, the mean is 4.84. If we are correct in thinking that the digits are drawn from a uniform supply, then we would wish, in another way of

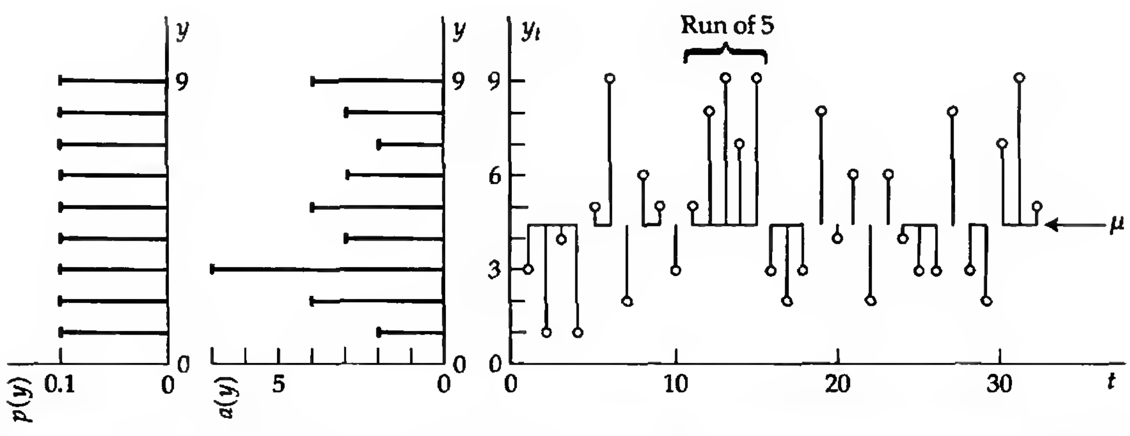


Fig. 17.2 **A finite sequence of digits y_t from the decimal representation of π, their occurrence histogram $a(y)$, and the probability $p(y)$ we choose to associate with the infinite sequence.**

expressing a consequence of this property, that the mean m would approach a limit μ equal to 4.5 as the number of digits N increased indefinitely, that is, that

$$\mu = \lim_{N\to\infty} m = 4.5.$$

This expectation could be made the basis of a (partial) test of the assumption of uniformity.

An alternative way of calculating the mean m is to count the number of nines and multiply by 9, multiply the number of eights by 8, and so on, and divide by N. Let $a(y)$ be the number of times the value y occurs; a histogram standing on end at the left of the sequence of digits in Fig. 17.2 represents $a(y)$, and it is tabulated below. So

$$m = \frac{a(1) + 2a(2) + \dots + 8a(8) + 9a(9)}{N} = \frac{1}{N}\sum_{y=0}^{9} ya(y).$$

Considering the situation as the number of terms N increases without limit, we have for the value μ which m approaches,

$$\begin{aligned}\mu &= \lim_{N\to\infty} m \\ &= \lim_{N\to\infty} \frac{1}{N}\sum_{y=0}^{9} ya(y) \\ &= \sum_{y=0}^{9} y \lim \frac{a(y)}{N}.\end{aligned}$$

There is no reason why we should not replace the limit of the sum by the sum of the separate limits. Now the quantity $a(y)/N$, which is here

y	0	1	2	3	4	5	6	7	8	9
$a(y)$	0	2	4	7	3	4	3	2	3	4
$a(y)/N$	0	.06	.12	.22	.09	.12	.09	.06	.09	.12

tabulated for our particular case, is the fractional number of times the value y occurs, and we believe that this is the same in the limit for all values of y from 0 to 9, namely 0.1. Calling this the probability $p(y)$ that a value y will occur, we write

$$p(y) \triangleq \lim_{N\to\infty} \frac{a(y)}{N}.$$

In Fig. 17.2 $p(y)$ is represented as a histogram standing on end. Since $\Sigma a(y) = N$, it is clear that $p(y)$ possesses the property $\Sigma p(y) = 1$ that is required of a probability distribution.

Finally, then, we can write

$$\mu = \sum_{y=0}^{9} yp(y);$$

or, in other words, the quantity μ, the mean in the limit of an indefinitely long sequence of values y_t, can be expressed, not only as the limit of m, the mean of N terms, but in a slightly more complicated way, as the first moment of the probability $p(y)$ governing the occurrence of each value y.

Similarly, the variance of the sequence $\{y_t\}$ approaches a limit which is the second moment of $p(y)$ about $y = \mu$; that is,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} (y_t - m)^2 = \sum_{y=0}^{9} (y - \mu)^2 p(y).$$

These results, which are well known in other connections, are restated here to form a link with the terminology to be employed.

Later, the sequence of random digits will be interpreted as a waveform, either by interpolation of smoothly varying intermediate values, or by imagining a regular sequence of impulses with strengths as given by the random digits. In the first case the spectrum will cut off at a frequency fixed by whatever the time interval between samples is taken to be, and in the second case the spectrum will run on to indefinitely high frequencies; but in each case the mean of the sequence is the same as the mean of the function.

FILTERING A RANDOM INPUT: EFFECT ON AMPLITUDE DISTRIBUTION

A random waveform generated from the sequence of random digits in either of the ways mentioned above may be passed through a filter, and the output may be examined to see what the effects have been. We think of a filter as a device that transmits different frequencies differently, and so one of the effects will be a change in the spectrum of the input waveform. But first we consider the probability distribution of amplitude and reach a remarkable conclusion. Since this conclusion is a basic consequence of summing random numbers, the discussion will be given without direct reference to filters, but what we are about to do is pass the random waveform through a filter whose response to an input $\{1\ 0\ 0\ 0\ \ldots\}$ is the 10-digit sequence $\{1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\}$.

Let us operate on our sequence of digits $\{y_t\}$ to form a new sequence $\{\eta_t\}$ according to the following rule: η_1 is the sum of the first 10 values of y_t, η_2 is the sum of the second to the eleventh inclusive, and so on; that is,

$$\eta_t = y_t + y_{t+1} + y_{t+2} + \ldots + y_{t+9}.$$

This sequence of numbers is graphed in Fig. 17.3, as far as the finite amount of data used in the previous graph allows. It is clear that the numbers are concentrated in the 30s, 40s, 50s, and 60s, and that there are no values falling outside this range. The histogram on the left shows the totals in each range. Just as in the case of the total occurrences of the 32 original digits, we believe that the fractional number of occurrences of each value y, where y ranges from 0 to 90, will narrow

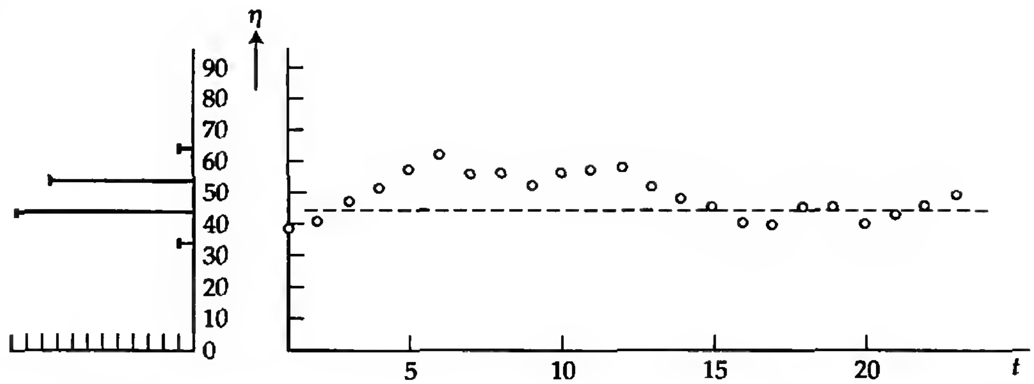


Fig. 17.3 **A sequence of digits into which correlation has been introduced between successive values, with corresponding spectral modification.**

in on some limit as the length of the sequence increases. We call this limit for each y the probability $p_{10}(y)$ that a sum of ten digits shall make a total y.

Since it was shown in Chapter 16 that the sum of two independent quantities having given probability distributions is distributed according to the convolution of the given distributions, we can immediately deduce $p_{10}(y)$, provided that successive digits in the decimal representation of π are independent. This raises a new and grave question.

Digression on independence. We could have avoided this question by postulating a sequence $\{y_t\}$ with the desired properties, without disclosing the actual values. However, since we have chosen to illustrate the subject by an actual example, a digression on independence is necessary; it is also appropriate, since the applicability of simple theory to actual data is an important matter.

It is quite possible that the digits may have equal relative frequencies of occurrence in the long run, yet at the same time each need not be independent of its predecessor. For example, the infinite decimal fraction 3.01234567890123456789..., which repeats in the way suggested, exhibits uniform frequency of occurrence of each digit, but successive digits are not independent.

A casual look at Fig. 17.2 does not reveal any obvious connections between one digit and the next; indeed, it has been established that in almost all decimals there is no connection, but it will certainly be in order here to apply some quantitative test such as the following. If y_{t+1} is truly independent of y_t, then it is as likely to be on the same side of 4.5 as it is to be on the opposite. Counting reveals that y_{t+1} crossed to the other side of 4.5 from y_t 15 times and remained on the same side 16 times. This is satisfactory; a stronger test is to examine the length of the runs above and below 4.5. Thus the first four digits 3, 1, 4, 1 are below, the next two are above, then we have one above, two below, and so on. A little farther there is a run of length 5 that is identified in Fig. 17.2. The lengths of the 16 runs are 4, 2, 1, 2, 1, 5, 3, 1, 1, 1, 1, 1, 3, 1, 2, 3. It is clear that the probability of a

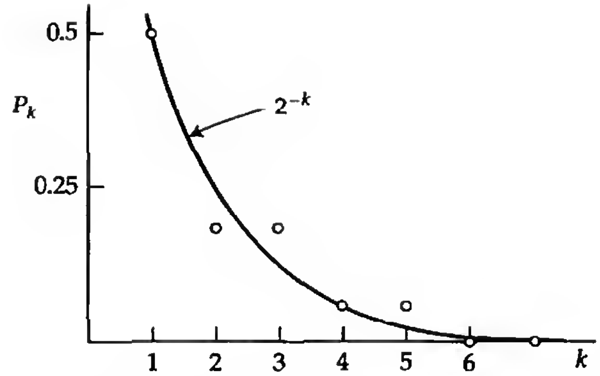


Fig. 17.4 The probability of a run of length k.

run of length 1 is the probability that a crossing that has just occurred will be immediately followed by a return, namely $\frac{1}{2}$. The probability of a run of length 2 is $\frac{1}{4}$, and in general the probability p_k of a run of length k is 2^{-k}.

A graph of $p_k = 2^{-k}$ is shown in Fig. 17.4, with points representing the actual total numbers of runs of different lengths in the example. One cannot fail to be struck by the agreement between theory and fact; but if we wished to pause and examine the significance of the agreement, it would be necessary to establish first the probability with which a run of length k occurs just n times out of 16. As an exercise the student might carry out a similar analysis of the length of runs of even digits and runs of odd digits.

As a more detailed check, construct a correlation diagram by plotting y_{t+1} against y_t as in Fig. 17.5. If this diagram revealed an elongation along the diago-

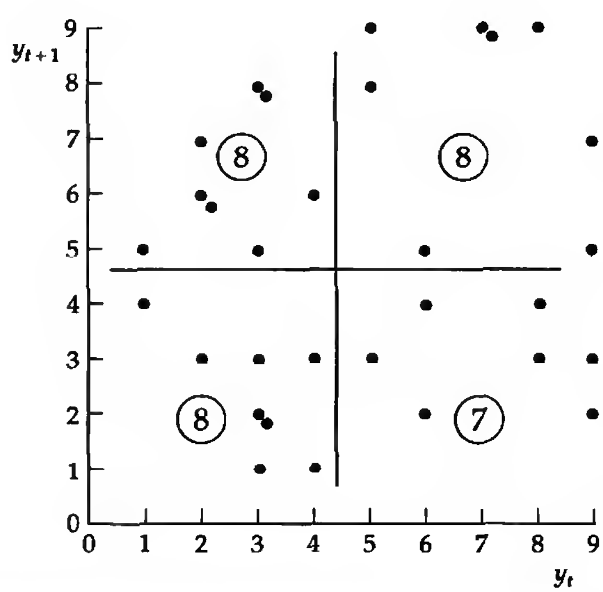


Fig. 17.5 Correlation diagram relating each digit y_{t+1} to its predecessor y_t.

nal through the first and third quadrants, it would mean a tendency of high values to be followed by high, and low values by low. This tendency could be evaluated numerically by calculating the covariance, or product moment about the centroid, $\Sigma(y_{t+1} - \langle y_{t+1}\rangle)(y_t - \langle y_t\rangle)$. If desired, the product-moment correlation coefficient could then be derived by dividing by the maximum possible value, but this entails much numerical work. Usually some other measure of correlation meets the demands of speed better; for example, we divide the diagram into quadrants meeting at the median point, and count the number of points in each quadrant. The excess in the first and third over those in the second and fourth, divided by the maximum possible value of that excess (the total number of points), provides an excellent correlation coefficient, which in the present case comes to 0.03. Here again we do not pause to look more closely into the significance of this result, but go on to note additional requirements imposed by independence.

Either of the correlation coefficients mentioned could have small values, while at the same time there could be some degree of dependence of y_{t+1} upon y_t. Figure 17.6 presents an artificial example. Therefore it should be noticed that correlation coefficients alone are not sufficient to guarantee independence. To *demonstrate* independence, it would be necessary to show that the correlation diagram occupied the square lattice with the expected degree of uniformity and, in addition, that the sequence of tracing out the diagram was free from order.

In these tests we have been concentrating on the possible dependence of one digit on its immediate predecessor; but since we propose to add together 10 consecutive digits, we should also look for such connections as could be revealed by constructing the eight further correlation diagrams of y_{t+2}, y_{t+3}, etc., up to y_{t+9}, against y_t.

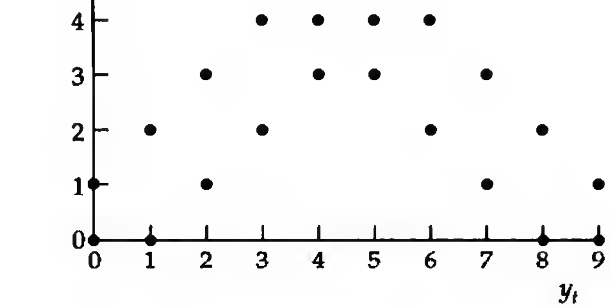


Fig. 17.6 **A correlation diagram showing dependence of y_{t+1} on y_t, which would not be revealed by correlation coefficients.**

At this point the writer will take the responsibility for asserting that successive digits are indeed independent.

This is an important step in the discussion of actual data, and the experimenter or observer must remember that it was taken, in case of later conflict, because one can never be absolutely sure that there is no underlying dependence. In effect, we have postulated a certain stochastic model, and we have found, from the partial tests so far made, no reason to doubt that it is a suitable choice as a basis for discussing the data in hand. We have also indicated the directions in which the theory of such tests, if elaborated, would lead. The responsibility for the degree of testing rests on the experimenter, who generally notes the absence of phenomena which, experience indicates, are most likely to complicate the situation.

The convolution relation. If the successive digits are independent, then the convolution relation holds, and $\{p_{10}(y)\}$ is the tenfold serial product of $\{p(y)\}$ with itself; that is,

$$\begin{aligned}\{p_{10}(y)\} &= \{\tfrac{1}{10}\ \tfrac{1}{10}\ \tfrac{1}{10}\ \tfrac{1}{10}\ \tfrac{1}{10}\ \tfrac{1}{10}\ \tfrac{1}{10}\ \tfrac{1}{10}\ \tfrac{1}{10}\ \tfrac{1}{10}\}^{*10}\\ &= 10^{-10}\{1\ \ 10\ \ 54\ \ 212\ \ 679\ \ldots\ (91\text{ terms})\ \ldots\ 54\ \ 10\ \ 1\}.\end{aligned}$$

The numerical values may be established directly by carrying out the successive convolutions. By the central-limit theorem we know that these values will be accurately given by a normal distribution with mean and variance each 10 times the mean 4.5 and variance 8.25, respectively, of the infinite sequence $\{y_t\}$. Hence the probability of occurrence of a value y in the infinite sequence $\{\eta_t\}$ is given approximately by

$$p_{10}(y) \approx \frac{1}{(2\pi \times 82.5)^{\frac{1}{2}}}\, e^{-(y-45)^2/(2\times 82.5)}.$$

Of course, this approximate expression assigns nonzero probabilities to values of y that are less than 0 and greater than 90, whereas we know from the method of construction of $\{\eta_t\}$ that such values are impossible. Therefore it would be possible to reveal that $p_{10}(y)$ was not in fact normal by examining a long-enough sequence. However, an extremely long sequence would be needed. Certainly the short sequence illustrated in Fig. 17.3 can be taken as indistinguishable from one derived from a strictly normal probability distribution.

The considerations that we have discussed reveal an important phenomenon. Because of the central-limit theorem, a sequence described by a probability distribution, when subjected to summing of successive terms, leads to a new sequence whose probability distribution is more nearly normal. Now, when a waveform is passed through a filter the output is a sum of successive input values, duly weighted, and so we may expect that normal probability distributions will be a common property of the amplitude of random signals emerging from filters.

The numerical process presented above as a running sum over ten terms is equivalent to passage through a filter with impulse response

{1 1 1 1 1 1 1 1 1 1}. What would be the result of passing the sequence $\{y_t\}$ through a filter with impulse response $\{g_0 g_1 g_2 \cdots\}$? As a concrete example, let

$$\{g\} = \{5\ 3\ 1\ 1\}.$$

Then the probability distribution of values of the output sequence is

$$\begin{aligned}
&\{1\ 0\ 0\ 0\ 0\ 1\ 0\ 0\ 0\ 0\ 1\ 0\ 0\ 0\ 0\ 1\ 0\ 0\ 0\ 0\ 1\ 0\ 0\ 0\ 0\ 1\ 0\ 0\ 0\ 0\\
&\qquad\qquad\qquad 1\ 0\ 0\ 0\ 0\ 1\ 0\ 0\ 0\ 0\ 1\ 0\ 0\ 0\ 0\ 1\}\\
&\quad * \{1\ 0\ 0\ 1\ 0\ 0\ 1\ 0\ 0\ 1\ 0\ 0\ 1\ 0\ 0\ 1\ 0\ 0\ 1\ 0\ 0\ 1\ 0\ 0\ 1\ 0\ 0\ 1\ 0\ 0\ 1\}\\
&\quad * \{1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\}\\
&\quad * \{1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\} \times 10^{-4}.\\
&\quad = 10^{-4}\{1\ 2\ 3\ 5\ 7\ 10\ 14\ 18\ 23\ 29\ \ldots\ (91\ \text{terms})\ \ldots\ 3\ 2\ 1\}.
\end{aligned}$$

In general, the probability distribution of the filter output is

$$\left[\frac{1}{g_0}p\left(\frac{y}{g_0}\right)\right] * \left[\frac{1}{g_1}p\left(\frac{y}{g_1}\right)\right] * \left[\frac{1}{g_2}p\left(\frac{y}{g_2}\right)\right] * \cdots.$$

This result shows in detail how *multiple* convolution enters into the amplitude distribution for only a *single* convolution connecting the input and output sequences. There is reason to think, therefore, that our sequence $\{\eta_t\}$ illustrates a widespread phenomenon, as regards its approximately normal amplitude distribution. However, this is not its only characteristic. We now have to consider the correlation that has been introduced between one term and the next. Successive terms are now clearly correlated, for each is the sum of ten numbers, nine of them common.

■

EFFECT ON AUTOCORRELATION

To examine the correlation between successive values of the sequence $\{\eta_t\}$ of Fig. 17.3, a plot of η_{t+1} versus η_t may be constructed as in Fig. 17.7. The points, which are numbered serially, reveal the strong correlation; evaluating the correlation as before by counting the points in the quadrants defined by the medians, we find a value of $(18 - 2)/20 = 0.8$. A similar calculation of the correlation of η_{t+2} gives 0.6, and so on, with decreasing values. These graphical calculations, which are quite speedy and visually informative, form a useful tool for practical data analysis. However, we now wish to calculate the correlation theoretically.

Consider the product-moment correlation coefficient. Let the sequence $\{\eta_t\}$ have N terms, let M_1 be the mean of the first $N - \tau$ terms, where $\tau = 0, 1, 2, \ldots,$ and let M_2 be the mean of the last $N - \tau$ terms. Then the correlation coefficient is defined by

$$\frac{\Sigma(\eta_{t+\tau} - M_2)(\eta_t - M_1)}{[\Sigma(\eta_{t+\tau} - M_2)^2\, \Sigma(\eta_t - M_1)^2]^{\frac{1}{2}}}$$

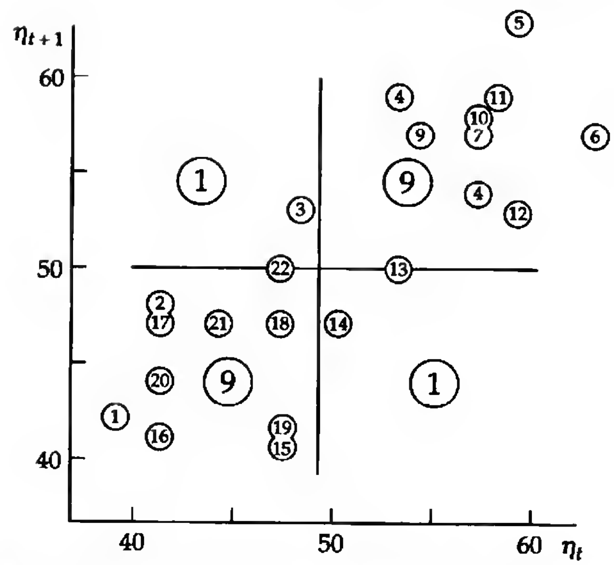


Fig. 17.7 **Correlation diagram relating successive pairs of terms of the sequence η_t. The correlation coefficient is $[9 + 9 - (1 + 1)] / 20 = 0.8$.**

where the summations run from $t = 1$ to $t = N - \tau$. There is a certain awkwardness in this expression because the means M_1 and M_2 are unequal and, furthermore, change with τ; it is nevertheless the precise expression for the product-moment correlation coefficient of a scatter of points (as in Fig. 17.7) about their centroid.

A simple and important result comes out if it is assumed at once that the correlation between one term and the next approaches a limit as the length of the sequence increases indefinitely, and if this limit is calculated.

For convenience in notation, we restate the problem in terms of sequences referred to their means in the limit. Thus

$$\mathbf{y}_t = y_t - \mu,$$

where

$$\mu = \lim_{N\to\infty} \frac{1}{N} \sum_1^N y_t$$

and

$$\mathbf{n}_t = \eta_t - \mu_1$$

where

$$\mu_1 = \lim_{N\to\infty} \frac{1}{N} \sum_1^N \eta_t.$$

Then

$$\lim_{N\to\infty} \frac{1}{N} \sum_1^N \mathbf{y}_t = \lim_{N\to\infty} \frac{1}{N} \sum_1^N \mathbf{n}_t = 0.$$

If the relation between these new centroid quantities is

$$\{\mathbf{n}_t\} = \{1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\} * \{\mathbf{y}_t\},$$

then the relationship is the same as in the original problem where ten consecutive values of y_t were added, except for end effects that will be insignificant for long sequences. For greater generality, write

$$\{\mathbf{n}_t\} = \{I_t\} * \{\mathbf{y}_t\},$$

where $\{I_t\}$ is a finite but more general sequence than the one previously considered. Without loss of generality we may, however, restrict $\{I_t\}$ so that

$$\sum I_t = 1.$$

Then the problem to be discussed is this: From a sequence of N uncorrelated random digits $\{y_t\}$ form a new sequence $\{\mathbf{y}_t\}$ by subtracting the mean, and form linear combinations $\{\mathbf{n}_t\}$ by taking the serial product with a sequence $\{I_t\}$. Find the correlation between members of $\{\mathbf{n}_t\}$, in the limit as N tends to infinity.

The product-moment correlation coefficient between $\{\mathbf{n}_t\}$ and itself displaced is

$$\frac{\sum \mathbf{n}_t \mathbf{n}_{t+\tau}}{[\sum \mathbf{n}_t^2 \sum \mathbf{n}_{t+\tau}^2]^{\frac{1}{2}}}$$

where the summations are suitably carried out. For convenience, we shall write this expression in the form

$$\frac{\mathbf{n}_t \star \mathbf{n}_t}{\text{normalizing factor}},$$

using the pentagram notation introduced in Chapter 3; it serves here to simplify the algebra. In the following argument we use the result that

$$(f * g) \star (f * g) = (f \star f) * (g \star g).$$

To derive this result, put f_r for the reverse of f; then by definition,

$$f \star g = f_r * g,$$

and

$$\begin{aligned}(f * g) \star (f * g) &= (f * g)_r * (f * g) = f_r * g_r * f * g = f_r * f * g_r * g \\ &= (f \star f) * (g \star g).\end{aligned}$$

Now let $C(\tau)$ be the limiting value of

$$\frac{\mathbf{n}_t \star \mathbf{n}_t}{\text{normalizing factor}}$$

for each displacement τ as $N \to \infty$. We shall refer to $C(\tau)$ as the autocorrelation function of the sequence $\{\mathbf{n}_t\}$, or of functions for which the sequence stands. Then

$$C(\tau) = \lim_{N \to \infty} \frac{\{\mathbf{n}_t\} \star \{\mathbf{n}_t\}}{\text{normalizing factor}}$$

$$= \lim_{N\to\infty} \frac{[\{I_t\} * \{\mathbf{y}_t\}] \star [\{I_t\} * \{\mathbf{y}_t\}]}{\text{normalizing factor}}$$
$$= \lim_{N\to\infty} \frac{[\{I_t\} \star \{I_t\}] * [\{\mathbf{y}_t\} \star \{\mathbf{y}_t\}]}{\text{normalizing factor}}$$
$$= \frac{[\{I_t\} \star \{I_t\}] * \lim_{N\to\infty}[\{\mathbf{y}_t\} \star \{\mathbf{y}_t\}]}{\text{normalizing factor}}.$$

But the members of $\{\mathbf{y}_t\}$ are uncorrelated; hence

$$\lim_{N\to\infty}\{\mathbf{y}_t\} \star \{\mathbf{y}_t\} \propto \{\ldots 0\ 0\ 1\ 0\ 0 \ldots\}$$

and
$$C(\tau) = \frac{\{I_t\} \star \{I_t\}}{\text{normalizing factor}}.$$

Since the normalizing factor was so chosen that $C(0) = 1$, it follows that the right-hand side is a normalized autocorrelation function itself.

Hence, finally, the autocorrelation of an initially uncorrelated sequence, after passage through a filter, is given by the autocorrelation of the impulse response of the filter.

■

EFFECT ON SPECTRUM

To consider the spectral effects of passing a noise waveform through a filter, it is well to reconsider the ways in which the sequence $\{x_t\}$ may be used to represent a waveform. Thus the string of numbers $\{x_t\}$ is fully equivalent to

$$\sum_j x_j \delta(t-j),$$

a succession of uniformly spaced impulses of strengths given by the numbers of the sequence, and also to the smooth curve

$$\sum_j x_j \operatorname{sinc}(t-j),$$

which passes through values given by the sequence. These two expressions have quite different spectra; the first has spectral components out to indefinitely high frequencies, whereas the spectrum of the second cuts off to zero. The following discussion will refer to the second case to avoid reference to two different kinds of spectrum, but the results will always be restatable in terms of an impulsive signal.

Spectrum of random input. Consider the waveform

$$V(t) = \sum_j \mathbf{y}_j \operatorname{sinc}(t-j)$$

corresponding to the sequence $\{y_t\}$, which has zero mean and equal probability of occurrence for the ten possible values that it can assume. What is the spectrum of this waveform?

Taking the Fourier transform directly,

$$V(t) \supset S(f)$$
$$= \sum_j y_j e^{-i2\pi jf} \Pi(f).$$

Looking at this result, we see that the transform is zero where $|f| > \frac{1}{2}$, and that for smaller values of f it can be regarded as the vector sum of a set of vectors of length y_j, oriented at progressively advancing angles $-2\pi jf$. The length of the resultant is the first thing to be calculated; its square will be given by multiplying the Fourier transform above by its complex conjugate. This is what we have previously called the power spectrum. It is given by

$$SS^* = \left(\sum_j y_j e^{-i2\pi jf}\right)\left(\sum_j y_j e^{i2\pi jf}\right)\Pi(f)$$
$$= [y_1^2 + y_2^2 + y_3^2 + \ldots + 2(y_1y_2 + y_2y_3 + y_3y_4 + \ldots)\cos 2\pi f$$
$$+ 2(y_1y_3 + y_2y_4 + y_3y_5 + \ldots)\cos 4\pi f + \ldots]\Pi(f)$$
$$= \left\{\sum_j y_j^2 + \sum_{\tau \neq 0}\left[\left(\sum_j y_j y_{j+\tau}\right)\cos 2\pi\tau f\right]\right\}\Pi(f).$$

The product moments $\sum y_j y_{j+\tau}$ are not in themselves zero, but the sequence $\{y_t\}$ is such that its autocorrelation function is zero for $\tau \neq 0$. By the definition, this means that $\sum y_j y_{j+\tau}$ becomes negligible relative to $\sum y_j^2$, for all τ except $\tau = 0$, as the number of terms in the sequence increases indefinitely, and hence the first term on the right-hand side becomes the dominant contribution. The important thing to notice is that this term does not contain the frequency f; in other words the power spectrum SS^* tends to become flat; it then cuts off at $f = \frac{1}{2}$.

Had we carried out the discussion in terms of the impulsive waveform $\sum x_j \delta(t - j)$, we would have found a power spectrum that was flat to indefinitely high frequencies.

A difficulty arises over assigning a numerical value to the power spectrum, for as is obvious the leading term $\sum y_j^2$ itself increases without limit as the length of the sequence increases. If a waveform is supposed to contain infinite energy (to convey power at a nonzero level for an infinite time), then there is no way of avoiding indefinitely large energy per unit bandwidth if all the energy is to be packed into a finite band.

It may be clear already that $\sum y_j^2$ is the total energy of the waveform, for

$$\int_{-\infty}^{\infty}\left[\sum y_j \operatorname{sinc}(t - j)\right]^2 dt = \sum y_j^2$$

and similarly,

$$\int_{-t}^{t} SS^*\, df = \sum y_j^2.$$

Of course, it follows from Rayleigh's theorem that these two integrals must be equal. One good way of avoiding the difficulty over infinite energy is to deal with waveforms of finite duration and to consider whether some quantity under discussion approaches a limit as the duration approaches infinity. The strength of the power spectrum happens not to, but the *shape* of the spectrum does.

From the autocorrelation theorem it follows that if the power spectrum is proportional to $\Pi(f)$, then the autocorrelation function will be sinc τ, and since this is equal to unity for $\tau = 0$, it is the normalized form; thus

$$C(\tau) = \operatorname{sinc} \tau.$$

To verify this result directly, we have

$$\begin{aligned}
C(\tau) &= \lim_{N\to\infty} \frac{V(t) \star V(t)}{\text{normalizing factor}} \\
&= \lim \frac{\left[\sum y_j \operatorname{sinc}(t-j)\right] \star \left[\sum y_j \operatorname{sinc}(t-j)\right]}{\text{normalizing factor}} \\
&= \lim \frac{\left\{\left[\sum y_j\, \delta(t-j)\right] * \operatorname{sinc} t\right\} \star \left\{\left[\sum y_j\, \delta(t-j)\right] * \operatorname{sinc} t\right\}}{\text{normalizing factor}} \\
&= \lim \frac{\left[\sum y_j\, \delta(t-j)\right] \star \left[\sum y_j\, \delta(t-j)\right] * \left[\operatorname{sinc} t \star \operatorname{sinc} t\right]}{\text{normalizing factor}} \\
&= \lim \frac{\left[\sum y_j^2\, \delta(\tau) + \sum_{\tau\neq 0}\left(\sum_j y_j y_{j+\tau}\right)\delta(t-\tau)\right] * \operatorname{sinc} \tau}{\text{normalizing factor}} \\
&= \frac{\left[\sum y_j^2\, \delta(t)\right] * \operatorname{sinc} \tau}{\text{normalizing factor}} \\
&= \frac{\sum y_j^2 \operatorname{sinc} \tau}{\text{normalizing factor}} \\
&= \operatorname{sinc} \tau.
\end{aligned}$$

Speaking in terms of the impulsive waveform $\Sigma y_j\, \delta(t-j)$ we would have found $C(\tau) = \delta^0(t)$, where $\delta^0(\tau)$ is the null function that is equal to unity at $\tau = 0$ and is equal to zero elsewhere, but some awkwardness would have arisen. For one thing, we wish $C(\tau)$ to be the Fourier transform of the power spectrum, but the transform of a null function is zero; therefore, as a consequence of introducing infinite-energy pulses in infinite numbers we would have to put up with a power spectrum that was flat out to infinite frequency, but equal to zero. It may

be noted, however, that the impulsive waveform proved convenient to introduce in the derivation given above.

Finally, the phase of the Fourier transform should be referred to, but it is sufficient to state here that the phase is equally likely to assume any value between 0 and 2π and to suggest as a problem the dependence of the phase on frequency, for a waveform of finite duration.

The output spectrum. When the waveform $V(t)$ is passed through a filter, the output waveform $W(t)$ is given by

$$W(t) = V(t) * I(t),$$

where $I(t)$ is the impulse response of the filter and is related to the sequence $\{I_t\}$, which was previously used to specify the filter, in the usual way, namely,

$$I(t) = \sum_j I_t \operatorname{sinc}(t - j).$$

It follows that the spectrum of $W(t)$ is given by the product of $S(f)$, the Fourier transform of $V(t)$, and $T(f)$, the transfer function of the filter. Thus

$$W(t) \supset S(f)T(f),$$

and the power spectrum of the output is given by

$$SS^*TT^*.$$

Since the power spectrum of the input waveform becomes flat out to its cutoff at $|f| = \frac{1}{2}$ as the duration of the waveform increases, we put $SS^* \propto \Pi(f)$, and hence the power spectrum of the output is proportional to

$$TT^*\Pi(f).$$

This is the Fourier transform of

$$[I(t) \star I(t)] * \operatorname{sinc} t,$$

a band-limited function whose representation as a sequence is

$$\{I_t\} \star \{I_t\}.$$

This is exactly in accord with the result found previously for the autocorrelation function $C(\tau)$ of the output, namely,

$$C(\tau) = \frac{\{I_t\} \star \{I_t\}}{\text{normalizing factor}}.$$

Evidently a key role is played by $\{I_t\} \star \{I_t\}$ and by its Fourier transform TT^*, which may be referred to as the power-transfer function of the filter.

If two filters $\{I_t\}$ and $\{J_t\}$ are cascaded, the key quantities become

$$(\{I_t\} * \{J_t\}) \star (\{I_t\} * \{J_t\})$$

and the over-all power-transfer function, which is the product of the separate functions. This observation permits handling the problem of random noise with any given power spectrum which is passed through a filter. We simply say that the given waveform is like noise with a flat spectrum that has already passed through one filter.

SOME NOISE RECORDS

Figure 17.8 shows a long sample of a noise waveform which was prepared by convolving the random digits published by Kendall and Smith (1940) with the sequence of binomial coefficients {1 5 10 10 5 1} and putting a smooth curve through the points so obtained. Since the mean value of digits distributed randomly from 0 to 9 is 4.5, the derived sequence has a mean of 144; the horizontal axis has been shown at this mean value. Let us consider the spectrum of the waveform.

It has already been shown that the autocorrelation coefficients in this situation will have values proportional to $\{I_t\} \star \{I_t\}$, where $\{I_t\}$ represents the impulse response of the filter; in the present case the numerical operation of convolution with {1 5 10 10 5 1} is equivalent to passage through a physical filter. Indeed the analogy is so close that one ordinarily speaks of numerical filtering. Hence the autocorrelation sequence for the series prepared as above, after subtraction of the mean, is

$$\{1\ 5\ 10\ 10\ 5\ 1\} \star \{1\ 5\ 10\ 10\ 5\ 1\} = \{1\ 10\ 45\ 120\ 210\ 252\ 210\ 120\ 45\ 10\ 1\}.$$

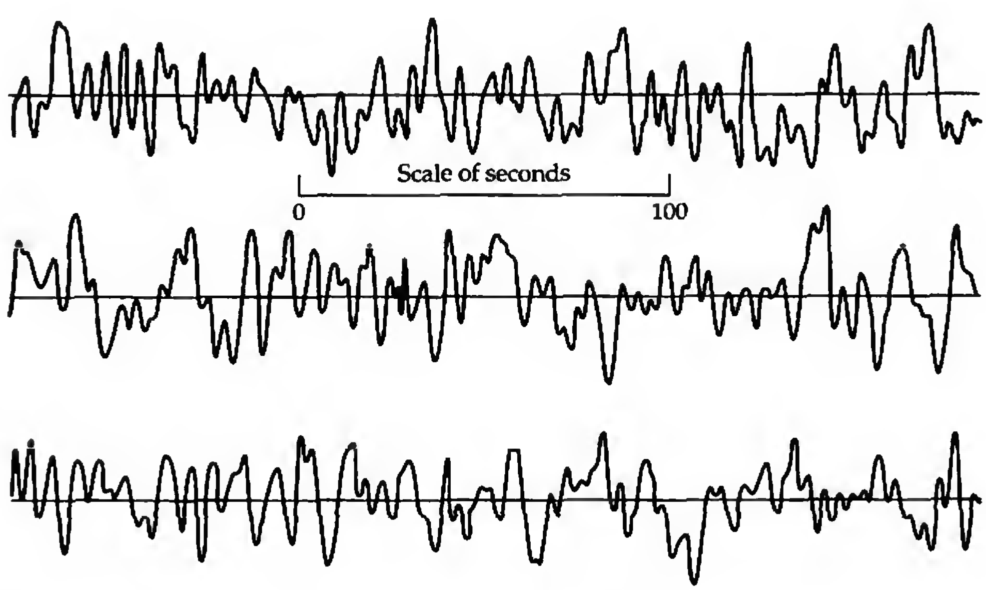


Fig. 17.8 **A sample of low-pass noise with a duration of 860 values.**

Now this binomial sequence may also be expressed as

$$\{1\ \ 1\}^{*10},$$

and hence it is to be expected from the central-limit theorem that the result will be approximately Gaussian. In fact, if values of the Gaussian expression $252\exp[-\pi(0.24n)^2]$ are calculated for integral values of n and compared with the binomial coefficients, we obtain:

n	0	1	2	3	4	5
Gaussian	252	210	122	49	14	3
Binomial	252	210	120	45	10	1

For compactness in the expressions that follow, the Gaussian expression will be substituted, but it is clear that the more precise results can be calculated just as easily.

Assume that the interpolation between points was carried out by convolution with a sinc function as in connection with the sampling theorem. Then the autocorrelation function of the smooth curve is proportional to

$$\{e^{-\pi(0.24\tau)^2}\text{III}(\tau)\} * \operatorname{sinc}\tau$$

and its power spectrum is proportional to the Fourier transform

$$\left\{\frac{1}{0.24}e^{-\pi(f/0.24)^2} * \text{III}(f)\right\}\Pi(f).$$

This last expression is approximately equal to

$$\frac{1}{0.24}e^{-\pi(f/0.24)^2}.$$

The degree of approximation is high, for the expression is already down to 10^{-5} at $f = 0.5$, where it must cut off abruptly to zero.

It follows that the noise record illustrated has an approximately Gaussian spectrum centered on zero frequency. The amplitude distribution may also be examined and will be found to be Gaussian to a good approximation. One need not hesitate, therefore, to accept the graph as an example of how Gaussian low-pass noise may be expected to behave. Many interesting features may be studied on the record, such as the general way in which the strength, rapidity, and character of the oscillations change as time elapses, and many quantitative questions are suggested, such as the rate and distribution of zero crossings, the distribution of maxima and minima, and so on.

Fig. 17.9 shows cases where the spectrum is now centered on a nonzero frequency, and the bandwidth is a fraction of the midfrequency. The top sample was constructed by drawing a smooth curve through a set of points $\{y_t\}$, each of which was calculated from the equation

$$y_t = ay_{t-1} + by_{t-2} + \epsilon_t,$$

where $a = 1.84$, $b = -0.9$, and ϵ_t was assigned values uniformly distributed at random between -0.5 and 0.5. The basis of this simple procedure is as follows. First,

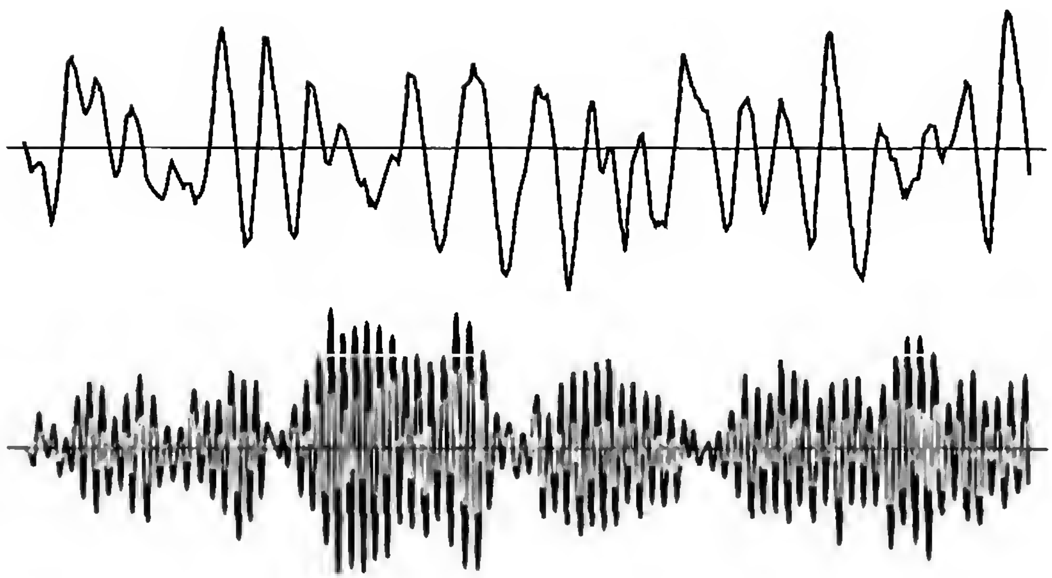


Fig. 17.9 Samples of bandpass noise.

consider the case in which the term ϵ_t is not included and where t is not restricted to integral values; thus let $y(t)$ be a function of t whose values at integral values of t are y_t. Then

$$y(t) = [a\,\delta(t-1) + b\,\delta(t-2)] * y(t).$$

This equation is obeyed by a damped oscillation

$$Ae^{-\delta t}\cos(\omega t + \alpha),$$

where A and α are an arbitrary amplitude and phase, respectively, and the angular frequency ω and the damping constant δ are related to the constants a and b by

$$a = 2 - \omega^2 - \delta$$
$$b = -(1 - \delta).$$

By inclusion of the random term we obtain the response of the damped resonator to random excitation which has a flat spectrum up to the cutoff frequency associated with the discrete spacing of successive terms.

The general character of bandpass noise is seen to consist of an oscillation centered on the midfrequency of the band and having an envelope amplitude that varies considerably. Close inspection would show that the phase of the oscillation drifts at a rate connected with the rate at which the amplitude of the envelope varies.

If the noise were passed through several damped resonators in tandem, each like the single one under discussion, the spectrum would approach a Gaussian distribution about its midfrequency. As it is, this example does not have a Gauss-

ian spectrum, but it does have a Gaussian amplitude distribution to a good degree of approximation. A question arises now that did not arise in connection with the low-pass noise sample: what is the distribution of the envelope amplitude?

ENVELOPE OF BANDPASS NOISE

To study the envelope of a noise waveform $y(t)$ we need to know its Hilbert transform $z(t)$. Then the amplitude of the envelope is defined, as explained in Chapter 13, by

$$\{[y(t)]^2 + [z(t)]^2\}^{\frac{1}{2}}.$$

Let the probability of finding $y(t)$ between y and $y + dy$ be $p_y(y)\,dy$, and let the probability of finding $z(t)$ between z and $z + dz$ be $p_z(z)\,dz$. In the case we are considering $p_y(y)$ is Gaussian; let

$$p_y(y) = ae^{-\pi(ay)^2}.$$

If, as will be shown in a moment, $z(t)$ has the same distribution, then

$$p_z(z) = ae^{-\pi(az)^2},$$

and if in addition z is independent of y, then the probability of finding $y(t)$ between y and $y + dy$, and simultaneously $z(t)$ between z and $z + dz$, is $p_y(y)p_z(z)\,dy\,dz$, where

$$p_y(y)p_z(z) = a^2e^{-\pi a^2(y^2+z^2)}.$$

Then on the (y, z)-plane the two-dimensional probability distribution can be written in terms of polar coordinates r and θ:

$$p_y(y)p_z(z)\,dy\,dz = a^2e^{-\pi a^2r^2}r\,d\theta\,dr.$$

Thus the probability $p(r)\,dr$ of finding the envelope amplitude $\{[y(t)]^2 + [z(t)]^2\}$ between r and $r + dr$ is given by

$$p(r)\,dr = a^2e^{-\pi a^2r^2}2\pi r\,dr,$$

or

$$p(r) = 2\pi ra^2e^{-\pi a^2r^2}.$$

This is the Rayleigh distribution, some of whose numerical parameters were recorded in Chapter 4. It is generally written, with reference to its mean-square abscissa $\langle r^2\rangle$, in the form

$$p(r) = \frac{2r}{\langle r^2\rangle}e^{-r^2/\langle r^2\rangle}.$$

Now the probability distribution of the Hilbert transform $z(t)$ has to be examined. The basic idea here is that $z(t)$ has statistics identical to those of $y(t)$; first, it has the same power spectrum and autocorrelation function. Regarding $y(t)$ as

the superposition of harmonic components in unrelated phases, we see that the process of Hilbert transformation, which shifts every Fourier component by a different amount, gives another but independent superposition of the same kind. On these assumptions it is thus found that the envelope is distributed according to a Rayleigh distribution.

DETECTION OF A NOISE WAVEFORM

If a noise waveform $y(t)$ is passed through a linear detector whose output is $|y(t)|$, then it is clear that the output waveform has the same Gaussian probability distribution as the input waveform, as regards positive outputs. Furthermore, the envelope of $|y(t)|$ is the same as the envelope of $y(t)$ and therefore has the same Rayleigh distribution.

If we use a square-law detector whose output is $[y(t)]^2$, the envelope V is the square of the envelope r of $y(t)$. Thus

$$V = r^2$$

and

$$dV = 2r\,dr.$$

Hence, since $p_V\,dV = p(r)\,dr$,

$$p_V = \frac{1}{\langle V\rangle}e^{-V/\langle V\rangle}.$$

This is a truncated exponential distribution.

MEASUREMENT OF NOISE POWER

Consider a noise waveform with power spectrum $R(f)$, and suppose that we are required to measure its strength. The signal might be one emerging from a bandpass amplifier connected to an antenna pointing at a wideband source of electromagnetic radiation; if the spectrum of the source is flat over the pass band of the amplifier, then $R(f)$ may be described as the power-transfer function of the reception filter.

Since the mean value of the signal is zero, it is passed through a detector to produce a unidirectional flow of current whose mean will be a measure of the strength. A square-law detector will be considered here, and we need to know that the output-power spectrum is given by

$$2R(f) \star R(f) + \left[\int_{-\infty}^{\infty} R(f)\,df\right]^2 \delta(f).$$

In this expression the first term enumerates the number of ways a difference frequency f (or sum frequency) can be found within the spectrum $R(f)$, and the second term represents the unidirectional component. The sum and difference frequencies arise from squaring a combination of harmonic components, as in the simple theory of detection of oscillatory signals.

To measure the mean value it is necessary to reduce the fluctuating part. Let this be done by passing the detected waveform through a linear smoothing filter whose power-transfer function is $S(f)$. Then its output has a spectrum

$$S(f)\,\{2R(f) \star R(f) + \left[\int_{-\infty}^{\infty} R(f)\,df\right]^2 \delta(f)\}.$$

The fluctuations are now reduced relative to the mean; the mean-square fluctuation is

$$2\int_{-\infty}^{\infty} S(f)[R(f) \star R(f)]\,df = 2R(f) \star R(f)\Big|_0 \int_{-\infty}^{\infty} S(f)\,df,$$

and the mean value is

$$[S(0)]^{\frac{1}{2}} \int_{-\infty}^{\infty} R(f)\,df.$$

The ratio of the root-mean-square fluctuation to the mean determines the limit to the precision that will be attainable in measuring the mean value, a limitation that is universally encountered in the theory of measurement. In this formulation

$$\frac{\text{rms fluctuation}}{\text{mean}} = \left[\frac{2R(f) \star R(f)|_0}{\int_{-\infty}^{\infty} R(f) \star R(f)\,df} \frac{\int_{-\infty}^{\infty} S(f)\,df}{S(0)}\right]^{\frac{1}{2}}$$
$$= \left(\frac{W_s}{\frac{1}{2}W_{R\star R}}\right)^{\frac{1}{2}},$$

where W_s is the equivalent width (Chapter 8) of the power-transfer spectrum of the smoothing filter, given by

$$W_s = \frac{\int_{-\infty}^{\infty} S(f)\,df}{S(0)},$$

and $W_{R\star R}$ is a quantity that has been described in Chapter 8 as the autocorrelation width of $R(f)$. It is the equivalent width of $R \star R$, the function that results when $R(f)$, including its negative frequency part, is substituted in the autocorrelation integral:

$$R \star R = \int_{-\infty}^{\infty} R(f')R(f + f')\,df'.$$

TABLE 17.1
Data for reception and smoothing filters†

Reception filter‡	$R(f)$	Δf
Rectangular pass band, $f_0 > \frac{1}{2}\Delta$	$\Pi\left(\frac{\lvert f\rvert - f_0}{\Delta}\right)$	Δ
Two separate rectangular pass bands	$\Pi\left(\frac{\lvert f\rvert - f_1}{\Delta_1}\right) + \Pi\left(\frac{\lvert f\rvert - f_2}{\Delta_2}\right)$	$\Delta_1 + \Delta_2$
Triangular pass band, $f_0 > \frac{1}{2}\Delta$	$\Lambda\left(\frac{\lvert f\rvert - f_0}{\frac{1}{2}\Delta}\right)$	$\frac{3}{4}\Delta$
Single tuned circuit, $f_0 \gg \Delta$	$\left[1 + \left(\frac{\lvert f\rvert - f_0}{\Delta}\right)^2\right]^{-1}$	$2\pi\Delta$
Two isolated tandem tuned circuits, $f_0 \gg \Delta$	$\left[1 + \left(\frac{\lvert f\rvert - f_0}{\Delta}\right)^2\right]^{-2}$	$\frac{4}{5}\pi\Delta$
Gaussian pass band§	$\exp\left[-\frac{1}{2}\left(\frac{\lvert f\rvert - f_0}{\Delta}\right)^2\right]$	$2\pi^{\frac{1}{2}}\Delta$
Smoothing filter	$S(f)$	τ
Takes running means over time T	$\operatorname{sinc}^2 Tf$	T
Single RC circuit	$[1 + (2\pi RCf)^2]^{-1}$	$2RC$
Two isolated tandem RC circuits	$[1 + (2\pi RCf)^2]^{-2}$	$4RC$
Critically damped RLC circuit	$\left[\left(\frac{R}{2L}\right)^2 + (2\pi f)^2\right]^{-2}$	$\frac{8L}{R}$
Rectangular pass band	$\Pi\left(\frac{f}{2f_0}\right)$	$\frac{1}{2f_0}$
Gaussian pass band	$\exp\left(-\frac{f^2}{2f_0^2}\right)$	$\frac{1}{(2\pi)^{\frac{1}{2}} f_0}$

†Bracewell (1962).
‡For bandwidths such that $R(0)$ is zero or negligible.
§The half-power bandwidth is $(8 \ln 2)^{\frac{1}{2}}\Delta$.

Thus

$$W_{R\star R} = \frac{\int_{-\infty}^{\infty} R(f) \star R(f)\, df}{R(f) \star R(f)|_0}.$$

It is convenient to write $\tau = \dfrac{1}{W_s}$ and $\Delta f = \frac{1}{2}W_{R\star R}$. Then

$$\frac{\text{rms fluctuation}}{\text{mean}} = \frac{1}{(\tau\, \Delta f)^{\frac{1}{2}}}.$$

In the simple situation where the reception filter receives uniformly in a band of width Δ and receives nothing outside this band, and the smoothing filter consists of a device that takes an average over a time interval T, then $\tau = T$ and $\Delta f = \Delta$. For other shapes of reception filter and other kinds of smoothing filter, the values of τ and Δf are presented in Table 17.1.

BIBLIOGRAPHY

Bracewell, R. N.: "Radio Astronomy Techniques," *Handbuch der Astrophysik*, vol. 54, pp. 50–129, Springer, Berlin, 1962.

Gardner, W. A.: "Statistical Spectral Analysis," Prentice-Hall, Englewood Cliffs, New Jersey, 1988.

Goodman, J. W.: "Introduction to Fourier Optics," 2nd ed., McGraw-Hill, New York, 1996.

Goodman, J. W.: "Statistical Optics," Wiley, New York, 1966.

Gray, R. M.: "Probability, Random Processes, and Ergodic Properties," Springer, New York, 1987.

Leon-Garcia, A.: "Probability and Random Processes for Electrical Engineering," Addison-Wesley, Reading, MA, 1989.

Papoulis, A.: "Systems and Transforms with Applications in Optics," McGraw-Hill, New York, 1968.

Parzen, E.: "Modern Probability Theory and Its Applications," Wiley, New York, 1960.

Ripley, B. D.: "Spatial Statistics," Wiley, New York, 1981.

Steward, E. G.: "Fourier Optics," Ellis Horwood, Chichester, U.K., 1987.

Swarup, G., A. R. Thompson, and R. N. Bracewell: "The Structure of Cygnus A," *Astrophys. J.*, vol. 238, pp. 305–309, 1963.

PROBLEMS

1. **Rule of thumb.** Before one can measure a signal-to-noise ratio, as, for example, from a record such as that of Fig. 17.1, it is first necessary to measure the strength of the noise. One way would be to read off values at closely spaced intervals and to calculate the root-mean-square value. However, according to a widely quoted rule of thumb, the root-mean-square value is one-fifth of the peak-to-peak value. Investigate this rule. ▷

2. **Accuracy of peak width.** From a record like that shown in Fig. 17.1, but in which the signal is of Gaussian form rather than double-peaked, the width of the signal at half-peak level is measured. Assuming that the signal-to-noise ratio is large, calculate the standard deviation of the measured width. How many times larger than the root-mean-square noise must the peak signal be if the width is to be measured with a standard deviation of 5 percent? ▷

3. A noise waveform has a Gaussian amplitude distribution and a certain power spectrum. What can be said about the amplitude distribution and power spectrum of its derivative?

4. **Dependence of upcross rate and peak rate on power spectrum.** (*a*) A sequence of random numbers with a Gaussian distribution is convolved with a sequence $\{I_t\}$. Let us say that successive members of the sequences are spaced 1 second apart. When one

member of the output sequence is negative and the following one is positive we say that a positive zero crossing, or upcross, has occurred. Show that the expected number of upcrosses per second ν is given by

$$\nu = \frac{1}{2\pi} \arccos \gamma_1,$$

where

$$\{\gamma_t\} = \frac{\{I_t\} \star \{I_t\}}{\text{normalizing factor}}$$

and the normalizing factor is chosen in the usual way so that $\gamma_0 = 1$.

(*b*) Show that knowledge of γ_1 is equivalent to knowing the central second difference of $\{\gamma_t\}$, or the central curvature of the autocorrelation function (ACF) of the smooth curve defined by $\{I_t\}$ when $\{I_t\}$ is a long sequence. Hence or otherwise show that the number of upcrosses per second ν is equal to the radius of gyration of the power spectrum $S(f)S^*(f)$, or

$$\nu^2 = \langle f^2 \rangle_{SS^*} = \frac{\int_{-\infty}^{\infty} f^2 S(f)S^*(f)\,df}{\int_{-\infty}^{\infty} S(f)S^*(f)\,df} = \frac{\text{central curvature of ACF}}{\text{central value of ACF}}$$

How would you interpret physically a situation where the integral in the numerator did not exist?

(*c*) Show that the number of maxima per second in a noise waveform whose power spectrum is $S(f)S^*(f)$ is equal to the number of upcrosses per second in a different waveform whose power spectrum is $f^2S(f)S^*(f)$. ▷

5. (*a*) The waveform shown in Fig. 17.8 has a Gaussian power spectrum $\exp(-f^2/2f_0^2)$. From the information given in the text regarding the construction of the figure, and assuming successive calculated values to be 1 second apart, determine f_0. ▷

(*b*) It is required to find f_0 for an observed waveform possessing a Gaussian power spectrum by counting the number of upcrosses per second. By actual count, estimate the value of the coefficient K in the relation $f_0 = K\nu$, where ν is the expected number of upcrosses per second. What is the precision of this estimate? ▷

6. Plot $\eta_{t+\tau}$ against η_t as in Fig. 17.7 for integral τ values up to 10, and graph the correlation against τ.

7. Let ρ be the product-moment correlation coefficient, and let ψ be the simple coefficient based on counting the dots in the quadrants of a scatter diagram as in Fig. 17.5. Under what circumstances will there be a one-to-one relationship between ρ and ψ? Show that

$$\rho = \sin\left(\frac{1}{2}\pi\psi\right)$$

is a useful conversion formula.

8. **Clipping noise peaks.** The wideband voltage induced in an antenna by radiofrequency radiation from outer space is expected to contain narrow-band signals caused by various interstellar molecules. Owing to severe technical difficulties associated with the faintness of the signal, computer processing is envisaged in which the received voltage is first clipped so as to assume a value of $+1$ whenever the voltage is positive and -1 when it is negative. It might seem that information would thereby be lost, but show that the autocorrelation function of the original signal is given by

$$\sin\left[\frac{1}{2}\pi C(\tau)\right],$$

where $C(\tau)$ is the autocorrelation function of the clipped signal, and hence that the presence of a spectral line in the radiation may be deduced by Fourier transformation. ▷

9. **Artificial random noise.** A signal $V(t)$ containing three closely spaced but incommensurable frequencies is given by

$$V(t) = \cos t + \cos et + \cos \pi t$$

Graph the signal, beginning at unit intervals of t, to gain an impression of the behavior for the first few cycles, and then quickly graph the envelope of the signal out to about $t = 150$. The envelope rises and falls in a way reminiscent of narrow-band noise, but the amplitude distribution cannot, of course, be Gaussian. Show that in fact the probability of finding the signal level between V and $V + dV$ is proportional to

$$(1 - V^2)^{-\frac{1}{2}}\Pi\left(\frac{V}{2}\right) * (2 - V^2)^{-\frac{1}{2}}\Pi\left(\frac{V}{2}\right) * (1 - V^2)^{-\frac{1}{2}}\Pi\left(\frac{V}{2}\right).$$

Show that the resemblance to Gaussian noise is so great that the amplitude distribution of a sample running on for hundreds of cycles would be difficult to distinguish from the real thing. ▷

10. **Mapping a radio source by coherence measurement.** The radio waves originating in our galaxy and other galaxies arrive at the earth with a strength which, for each direction of arrival, is specified by a temperature T (a convenient quantity which, in the frequency range concerned, is directly proportional to power per unit frequency interval). Two identical antennas immersed in the field are equipped with identical narrow-band amplifiers. The amplifier output voltages, which are given by the (complex) time-dependent phasors $V_1(t)$ and $V_2(t)$ are transmitted by identical transmission lines to a nonlinear circuit which measures the complex quantity Γ defined by

$$\Gamma = \frac{\langle V_1(t)V_2^*(t)\rangle}{\langle V_1(t)V_1^*(t)\rangle}.$$

The quantity Γ is known as the complex degree of coherence of the field as a function of spacing between points occupied by the two antennas. If attention is confined to a small solid angle, for example, by the use of directional antennas, show that the sky source can be mapped using

$$\frac{\overline{\overline{T}}}{\overline{\overline{T}}|_0} = \Gamma,$$

where $\overline{T}$ is the Fourier transform of the wanted temperature map T. State the nature of the independent variables on which $\overline{T}$ and Γ depend. *Hint*: One approach to this problem is to deal in terms of a transverse field component $F(t)$ on a remote plane perpendicular to the line of sight (the plane being supposed to be less remote than the sources of energy). The quantity T may be introduced as a statistic of this field expressible by $T \propto \langle F(t)F^*(t)\rangle$. ▷

11. The power-transfer function of an amplifier, as a function of frequency f, is $\exp(-f^2/2B^2)$. The amplifier is followed by a square-law detector, a low-pass filter whose power-transfer function is $\exp(-f^2/2b^2)$, and a recording ammeter. As an over-all-gain calibration procedure, a signal from a well-designed noise generator is applied to the amplifier and the resulting deflection is read from the record. The noise power from the diode has a spectrum that is uniform from a low frequency that is negligible compared with B to a high frequency much greater than B and its level is well standardized. Show that the rms fluctuation on the record is K times the deflection, where

$$K = \left(\frac{2^{\frac{1}{2}}b}{B}\right)^{\frac{1}{2}}.$$

Suppose that $B = 10^4$ and $b = 10^2$ and that the precision requirements are such that gain changes of 1 percent should be detectable. Would it take more or less than 1 second to establish that the gain had changed by 1 percent? How would you investigate gain fluctuations taking place with characteristic times of the order of $\frac{1}{2}$ second? ▷

12. A voltage $V_1(t)$ is applied to a band-pass amplifier whose impulse response is $I(t)$, squared by a detector, and smoothed by a low-pass filter whose impulse response is $J(t)$. Show that the output voltage $V_2(t)$ is given by

$$V_2(t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} V_1(t-\tau_1)K(\tau_1,\tau_2)V_1(t-\tau_2)\,d\tau_1\,d\tau_2,$$

where

$$K(\tau_1,\tau_2) = \int_{-\infty}^{\infty} I(\tau_1-\tau)J(\tau)I(\tau_2-\tau)\,d\tau.$$

13. The spectrum of a noise waveform with Gaussian amplitude distribution extends up to but not beyond a frequency f_c. The critical rate of sampling given by the sampling theorem is thus $2f_c$. Show that the samples are not in general independent, and deduce under what circumstances the samples are uncorrelated. The precision with which noise power may be measured is expressible in terms of the square root of the number of independent measures. Show that the rate of occurrence of effectively independent values is connected with the autocorrelation width of the power spectrum rather than with f_c. ▷

14. **Surface tolerance of a reflector antenna.** Spherical wavefronts diverging from a point source at the focus of a perfect paraboloidal reflector become plane after reflection, but because of irregularities in construction permitted by the design, the wavefront emerging from a certain antenna is corrugated in a way resembling Fig. 17.8. Let y be the de-

parture of the wavefront from the plane of best fit, and let λ be the operating wavelength. Show that, at a distant receiving point on the normal to the plane, the received power falls below what could be received from a perfect reflector by a factor

$$\frac{1}{1 + \mathcal{D}\langle\delta^2\rangle}$$

where $\delta = 2\pi y/\lambda$. The angular brackets signify averaging over the antenna aperture, and $\mathcal{D}$ is a design parameter which is equal to unity if the intended field intensity is the same at all points of the emerging wavefront. ▷

15. **Random antenna array.** A number N of identical antennas, all pointing at the zenith, are arranged at random on an area of ground G that is so large that the possibility of antennas overlapping does not have to be considered. The antennas are connected to a single pair of terminals through equal lengths of transmission line in such a way that signals from a distant point source in the direction of the zenith arrive at the terminal pair in the same phase. To explore the field-radiation pattern of such a "random array," we move the point source about on the surface of a sphere that is large compared with G and is centered on the centroid of G, and measure the terminal voltage in amplitude and phase. The voltage at the terminals falls off as the point source moves away from the zenith, and for each source direction we consider the ensemble-average voltage, that is, the average of the voltage phasors produced by all possible arrangements of the N antennas within G. Show that the ensemble-average voltage is proportional to the voltage that would be received if G were completely filled with antennas, and that the field-radiation pattern of the random array is therefore the same, on the average, as that of one huge antenna occupying the whole area G. ▷

16. A peak voltmeter contains a diode in series with a resistor and capacitor in parallel. An alternating voltage applied across this circuit forces charge into the capacitor during brief intervals at the voltage peaks at an average rate that just balances the leakage of charge from the capacitor through its shunt resistor. A meter across the capacitor indicates the peak voltage of the applied waveform. Investigate the suitability of such an instrument for measuring the rms voltage of a noise waveform.

17. **Turbulent trail.** The trail left in the atmosphere by a meteor soon develops many wiggles, and the effects of the twisting of the trail are observable by radar, and also by photography when the persistence of visibility is sufficient. One of the physical causes originally proposed to explain the distortion was random turbulent motion of the atmosphere, the theory of which automatically generates mean-square quantities. To test this theory, the trail is supposed to be projected onto a plane, and its slope with respect to, and its departure from, a straight line of best fit are considered. It is deduced that the number of upcrosses per unit length ν is given by

$$2\pi\nu = \frac{\text{root-mean-square slope}}{\text{root-mean-square departure}}.$$

Investigate this result.

18. **Random noise and linear detector.** A zero-mean random noise voltage waveform $v(t)$ with standard deviation σ is passed through a linear detector so that the output is zero

when $v(t)$ is negative but is unchanged when $v(t)$ is positive. The output then passes through a low-pass filter. Let us say that the pass band of the filter is so related to the spectrum of $v(t)$ that the filter output is the average of five effectively independent values of the detector output. The filter output fluctuates about a mean voltage m with a standard deviation σ_f.

(*a*) Show that

$$m = (2/\pi)^{\frac{1}{2}}\sigma$$

$$\sigma_f = \left(\frac{1 - 2/\pi}{5}\right)^{\frac{1}{2}}\sigma.$$

(*b*) If the waveform $v(t)$ has a resonance spectrum with a 3-decibel bandwidth of 1000 hertz centered on 6000 hertz, and if the filter is a digital counter whose output accurately averages the detector output for 5 milliseconds, show that

$$\sigma_f = \left(\frac{1 - 2/\pi}{5\pi}\right)^{\frac{1}{2}}\sigma.$$

18

Heat Conduction and Diffusion

The subject of heat conduction holds a special place in Fourier theory, because it was Fourier's work on problems connected with heat that led him to expound his ideas in his "Théorie Analytique de la Chaleur," published in 1822. In this chapter it is shown how Fourier analysis enters into the study of heat flow, so that the background of ideas acquired thus far may be directly applied to this field. For simplicity only one-dimensional flow of heat is discussed, but this is sufficient for our purposes. Furthermore, it will be clear how the Fourier transforms in two and three dimensions would be required in the more general geometrical situation. For general background see Jakob (1957).

It should not be thought that the ideas of this chapter are confined to heat conduction; they refer basically to diffusion and to behavior in general that is described by the diffusion equation. Thus the kind of physics involved is the same as that governing the diffusion of electrons and holes in semiconductors, the diffusion of electrons in a plasma, or ionized meteor trails in the ionosphere. But for convenience the text is written in terms of heat conduction, with some reference to electric phenomena in cables.

ONE-DIMENSIONAL DIFFUSION

The diffusion equation

$$\nabla^2 V = \frac{1}{K}\frac{\partial V}{\partial t}$$

or, in its one-dimensional form,

$$\frac{\partial^2 V}{\partial x^2} = \frac{1}{K}\frac{\partial V}{\partial t},$$

is well known as describing the phenomena of heat conduction and diffusion. We shall confine attention to the one-dimensional equation, which suffices for the present discussion of diffusion, and only differs from the more general equation in being less complex geometrically.

If $V(x)$ is the temperature in a bar in which the heat can flow only in the $\pm x$ directions, then there will be flow of heat only where there is a gradient of temperature $\partial V/\partial x$. The amount of heat per second, I, which can be urged along the bar, is proportional to the temperature gradient and is inversely proportional to the thermal resistance r of the material of the bar per unit length. Thus

$$I = -\frac{1}{r}\frac{\partial V}{\partial x}.$$

The conducted heat, having flowed down the temperature gradient to places where temperature is lower, produces a temperature rise, which is proportional to the amount of heat accumulating per unit length. The property of the material which describes the amount by which the temperature will rise is the thermal capacitance[1] per unit length, c; the temperature rise is inversely proportional to c. The amount of heat accumulated in unit length per second is the difference between what flows in and what flows out, namely, $\partial I/\partial x$. Thus

$$\frac{\partial V}{\partial t} = -\frac{1}{c}\frac{\partial I}{\partial x}.$$

This equation allows for a falling temperature where the received heat is insufficient to supply what must flow out, that is, where $\partial I/\partial x$ is positive, or the rate of change of temperature gradient $\partial^2 V/\partial x^2$ is negative. It will be seen that this tends to make $V(x)$ smoother as time progresses, and as we possess quantitative measures of smoothness we can give a definite meaning to this statement.

A basic difference between diffusion and wave propagation shows up at this point. The smoothness of a stretched string or a water surface in the presence of waves described by the (nondissipative) one-dimensional wave equation

$$\frac{\partial^2 V}{\partial x^2} = \frac{1}{v^2}\frac{\partial^2 V}{\partial t^2},$$

where v is the wave velocity, cannot increase as time elapses, as a general property, for the waves described by substituting $-t$ for t are also solutions of the wave equation. Under any reasonable definition, therefore, the smoothness of a string must remain the same as time elapses. In diffusion, on the other hand, the phenomena are not reversible; a sequence of distributions V would not be compatible with nature if considered in reverse order.

---

[1]For bars of unit cross section, r and c may be translated into conventional quantities by replacing r^{-1} by thermal conductivity and c by volumetric specific heat ρs, where ρ is density and s specific heat. The units of r and c are the thermal ohm, or thohm, per meter and the thermal farad, or tharad, per meter. The author is indebted to W. Shockley for this information. Obviously, V and I are to be measured in tholts and thamps.

From the two equations describing one-dimensional heat flow we obtain, by elimination of $\partial I/\partial x$,

$$\frac{\partial^2 V}{\partial x^2} = rc\frac{\partial V}{\partial t}.$$

By appropriate redefinition of symbols one can establish the same partial differential equation wherever the gradient of a physical quantity V is the driving agent causing, in linear proportion, transfer of the same stuff whose accumulation linearly raises V. Examples of such quantities V are the concentration of a chemical in solution, the partial pressure of a gaseous constituent of an atmosphere, the temperature in a medium, the potential difference between the conductors of a submarine cable, and the concentration of holes or electrons in a semiconductor. We choose the cable as our second example because it permits representation by a circuit diagram. The conductors have a combined resistance r and an electrical capacitance c per unit length. The potential difference is $V(x)$, and the current is $I(x)$ in one conductor and $-I(x)$ in the other. The electric charge stored per unit length is cV. Any inductive reactance is negligible compared to the resistance of the conductors, and conduction currents in the medium separating the conductors are negligible in comparison with the displacement current. Then the equations

$$I = -\frac{1}{r}\frac{\partial V}{\partial x}$$

$$\frac{\partial V}{\partial t} = -\frac{1}{c}\frac{\partial I}{\partial x}$$

will apply to the cable, and a length dx of the cable may be represented by circuit (*a*) of Fig. 18.1. A finite length of cable may be regarded as a chain (*b*) of infinitesimal elements (*a*). The diffusion of heat along a poker whose end has been thrust in the fire is analogous to the penetration of charge from a battery into a cable (Fig. 18.2). The quantity $q = cV$ is the linear charge density, or, in the thermal case, the quantity of heat per unit length.

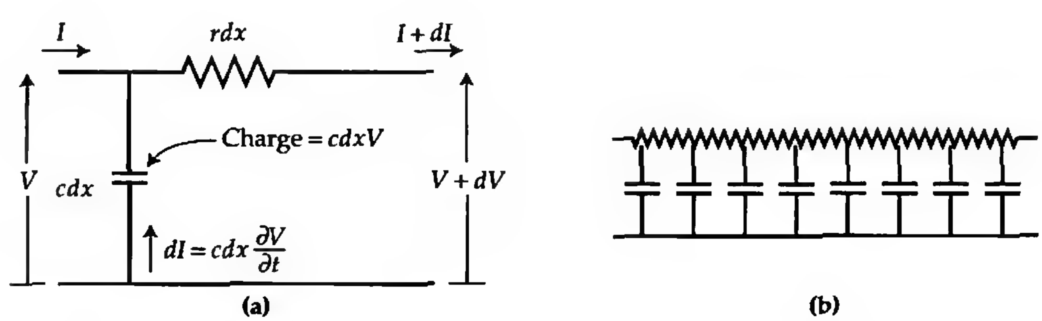


Fig. 18.1 **Circuit representation of an electrical cable that has only resistance and capacitance.**

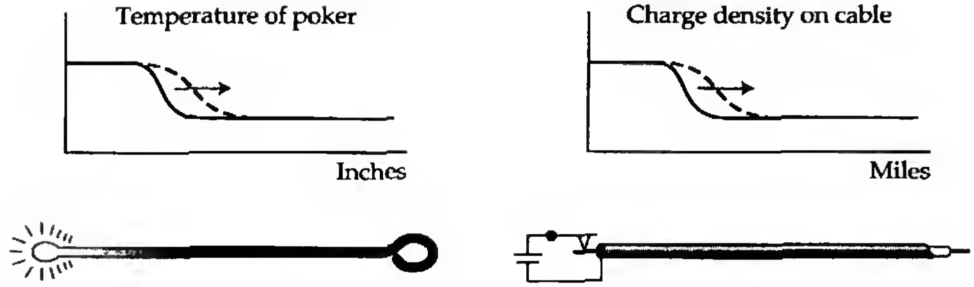


Fig. 18.2 **Diffusion of heat along a poker and of electric charge into a cable.**

The irreversibility of time in heat-conduction phenomena can be associated with the change of entropy, which increases at the rate

$$\int \frac{c(\partial V/\partial t)}{V}\,dx.$$

In the electrical case the irreversibility is associated with the dissipation of energy at a rate

$$\int \tfrac{1}{2}\,rI^2\,dx.$$

In the conduction case entropy increases until an isothermal state is reached and the flow of heat ceases; in the electrical case the final state of zero energy dissipation occurs when each conductor reaches constant potential and the flow of current ceases.

Does this mean that increase of entropy during heat flow corresponds to conversion of electrical to thermal energy? No, for the requirement on electrical dissipation holds point by point whereas entropy may decrease at some points while increasing at others, only the integral over all points being subject to the requirement of monotonic increase.

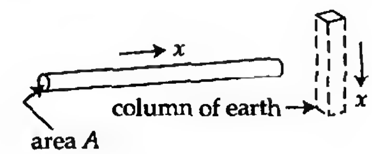
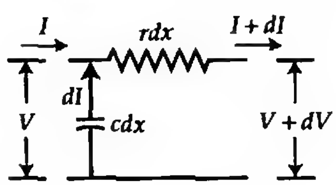
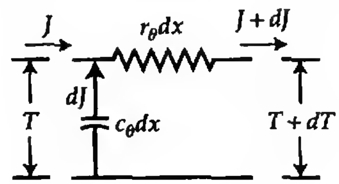
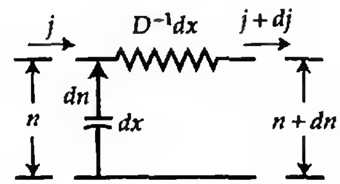
It is well known that resistance-capacitance networks are nonoscillatory, a reason being that in the absence of inductance we do not have the two different kinds of energy storage element upon which ordinary network oscillations depend. One or more pulsations are not excluded from the natural behavior, but if in the beginning (or at any time) there are no negative potentials, then none can subsequently develop, as they would in an ordinary oscillatory circuit. Of course, a nonisothermal system of heat conductors left to itself never develops temperatures lower than the original minimum.

The simple diffusion equation arises in a variety of fields of physics, chemistry, and engineering and the same transform approach can be applied to all; diffusion of electrons and holes in semiconductors is one example. Transference of a basic theory to a different field is impeded by the nonuniformity of symbols and the choice of parameters to which the symbols are assigned. Such relationships are illustrated in Table 18.1.

TABLE 18.1

Analogies between diffusion of heat in thermal conductors, diffusion of electric current carriers in semiconductors, and propagation of electric signals in a cable without inductance or leakage conductance.

Equation (1) gives the voltage, temperature, and particle density gradients. Equation (2) gives the current, heat flow, and particle flow gradients. Eliminating the current I (or J or j) leads to the familiar diffusion equations (3). Equation (1) in its three versions states that flow is proportional to gradient. Equation (2), which expresses the Kirchhoff law $\Sigma i = 0$ at the input node, corresponds to the continuity equation of fluid dynamics.

Electric	Thermal	Semiconductor	
			
Voltage V (volts)	Temperature T (kelvins)	Carrier density n (particles cm^{-3})	
Current I (amps)	Heat flow J (watts)	Flow density j (particles $cm^{-2}\,s^{-1}$)	
Voltage gradient $\partial V/\partial x$ (volt m^{-1})	Temperature gradient $\partial T/\partial x$ (K m^{-1})	Density gradient $\partial n/\partial x$ (cm^{-4})	
Resistance/m r ($\Omega\ m^{-1}$)	Thermal res./m r_θ (thohm m^{-1})	(Diffusion const.)$^{-1}$ D^{-1} (s cm^{-2})	
Capacitance/m c (F m^{-1})	Thermal capacitance c_θ	Unity and dimensionless	
			
$rI = -\partial V/\partial x$	$r_\theta J = -\partial T/\partial x$	$D^{-1} j = -\partial n/\partial x$	(1)
$c\partial V/\partial t = -\partial I/\partial x$	$c_\theta \partial T/\partial t = -\partial J/\partial x$	$\partial n/\partial t = -\partial j/\partial x$	(2)
$\dfrac{\partial^2 V}{\partial x^2} = rc\dfrac{\partial V}{\partial t}$	$\dfrac{\partial^2 T}{\partial x^2} = r_\theta c_\theta \dfrac{\partial T}{\partial t}$	$\dfrac{\partial^2 n}{\partial^2 x} = \dfrac{1}{D}\dfrac{\partial V}{\partial t}$	(3)

When an electric cable is modeled by equations containing only resistance and capacitance per unit length, modification is required if dielectric loss becomes significant; in that case conductivity g per unit length may be taken into account by inserting in the equivalent circuit a conductance in parallel with the capacitance. When thermal diffusion along a conductor is affected by lateral heat loss, the same circuit modification applies. When plasma devices are excited electrically at such high frequencies that the kinetic energy of the electrons begins to influence behavior, an inductance is to be placed in parallel with the capacitance. If plasma density becomes so high that collisions between electrons and plasma particles cannot be ignored then that inductance acquires a resistance in series with it.

The electric circuit analogy, which is purely mathematical, standing as it does for a set of equations based on Kirchhoff's laws, illuminates abstruse equations in unfamiliar fields and confers the power of applying physical intuition that one may possess from experience in one field to a new discipline.

GAUSSIAN DIFFUSION FROM A POINT

Two special cases will now be considered. Let the initial distribution of V at $t = 0$ be given by

$$A\,\delta(x),$$

where A is a constant. Then it may be verified by substitution that for subsequent values of t,

$$V_t(x) = A\left(\frac{rc}{4\pi t}\right)^{\frac{1}{2}} e^{-x^2rc/4t}$$

and

$$I_t(x) = A\frac{rcx}{2t}\left(\frac{rc}{4\pi t}\right)^{\frac{1}{2}} e^{-x^2rc/4t}.$$

In other words the distribution V is a Gaussian one which broadens as $t^{\frac{1}{2}}$ and diminishes in amplitude as $t^{-\frac{1}{2}}$ so that the area under it remains constant. The flow I is representable by two Rayleigh distributions which broaden as $t^{\frac{1}{2}}$ and diminish as t^{-1}. The distributions of V and I are shown in Fig. 18.3. At any nonzero value of x the values of both V and I pass through maxima before subsiding to zero.

At any instant after $t = 0$ the total stored energy in the electrical case is given by

$$\int_{-\infty}^{\infty} \tfrac{1}{2}cV^2\,dx = \tfrac{1}{2}c\frac{A^2rc}{4\pi t}\int_{-\infty}^{\infty} e^{-2x^2rc/4t}\,dx = \frac{A^2rc^2}{8\pi t}\left(\frac{4\pi t}{2rc}\right)^{\frac{1}{2}}$$

$$= A^2\left(\frac{rc^3}{32\pi rt}\right)^{\frac{1}{2}}.$$

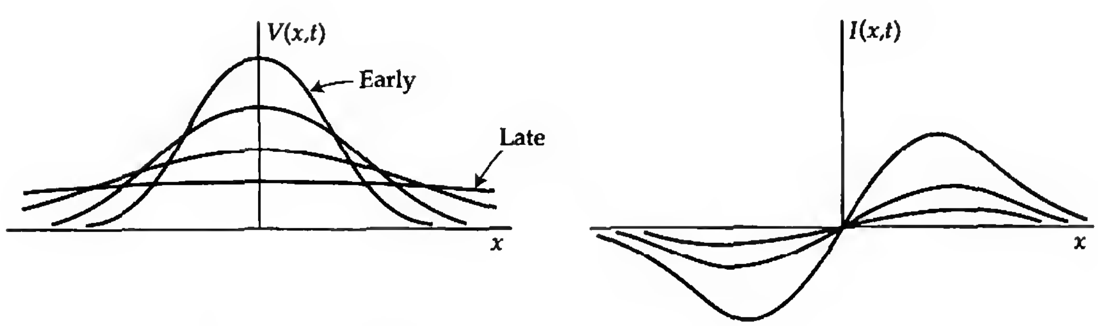


Fig. 18.3 The consequences of a local injection of heat.

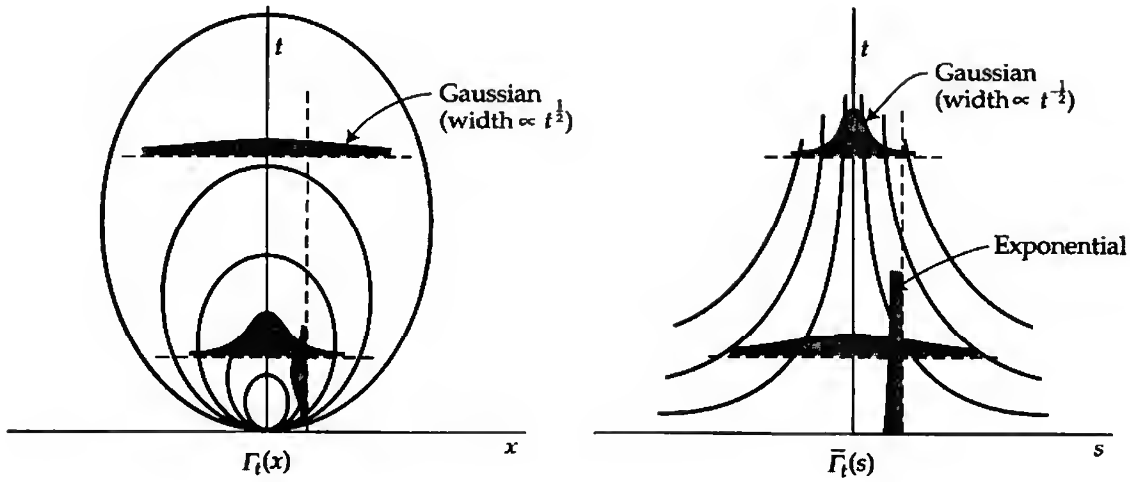


Fig. 18.4 **The Green's function for the diffusion equation, $\Gamma_t(x)$, and its Fourier transform.**

Knowing that a concentrated deposit of heat diffuses away in a Gaussian manner as described by

$$\Gamma_t(x) = \left(\frac{rc}{4\pi t}\right)^{\frac{1}{2}} e^{-x^2 rc/4t},$$

and bearing in mind the linearity of the diffusion equation, which permits the superposition of different solutions, we may express the diffusion of an arbitrary initial distribution $V_0(x)$ in the form of a convolution integral over x:

$$V_t(x) = \Gamma_t(x) * V_0(x).$$

At any positive value of t we can say that the distribution V is the convolution of the initial distribution with the Gaussian distribution to which a point concentration would have died away in the time t. The function $\Gamma_t(x)$ would ordinarily be referred to as the Green's function for the diffusion equation. It is illustrated in Fig. 18.4.

■

DIFFUSION OF A SPATIAL SINUSOID

Special interest now attaches to the Fourier transforms $\overline{\Gamma}_t(s)$ of the Gaussian diffusion functions $\Gamma_t(x)$, for Fourier theory can now be introduced via the convolution theorem. We know that

$$\overline{\Gamma}_t(s) = e^{-4\pi^2 ts^2/rc},$$

and note that $\overline{\Gamma}_t(0) = 1$ for all t just as $\Gamma_t(x)$ has the same area for all t. It follows from the convolution theorem that

$$\overline{V}_t(s) = e^{-4\pi^2 ts^2/rc}\,\overline{V}_0(s),$$

which can be interpreted as follows. Suppose $V_t(x)$ to be resolved into spatial Fourier components having the form ${}^{\cos}_{\sin}2\pi sx$. Then each such component decays exponentially with time with a time constant

$$\frac{rc}{4\pi^2 s^2};$$

that is, the higher the spatial frequency s, the more rapid the decay, as would be expected from the higher temperature gradients associated with the shorter spatial wavelength. For $s = 0$, that is, for a uniform distribution, the time constant is infinite, as, of course, it must be since there is no gradient to cause any heat flow.

We can now picture the diffusion of any initial distribution as follows. Subject the distribution to a low-pass filtering operation in which the filter characteristic is Gaussian and gives no phase change. According to the time which is to elapse, adjust the filter so that its width shrinks in proportion to $t^{-\frac{1}{2}}$. In effect we are saying that

$$^{\cos}_{\sin}2\pi sx\, e^{-4\pi^2 s^2 t/rc}$$

represents natural behavior, that (co)sinusoidal distributions remain (co)sinusoidal, simply dying away without changes of spatial frequency and without x shift, and that, therefore, Fourier analysis leads to a simple description of the diffusion phenomenon. The behavior of distributions depending (co)sinusoidally on x can be derived directly from

$$\frac{\partial^2 V}{\partial x^2} = rc\frac{\partial V}{\partial t}$$

by noting that partial differentiation with respect to x will be equivalent to multiplication by $i2\pi s$. Then

$$(i2\pi s)^2\mathbf{V} = rc\frac{d\mathbf{V}}{dt},$$

where $\mathbf{V}$ is the phasor such that the real part of $\mathbf{V} \exp i2\pi sx$ is the actual distribution. Thus

$$\mathbf{V} = Ae^{i\Phi}e^{-4\pi^2 s^2 t/rc},$$

where $A \exp i\Phi$ is an arbitrary complex constant and

$$\begin{aligned} V &= \operatorname{Re}\,[Ae^{i\Phi}e^{-4\pi^2 s^2 t/rc}e^{i2\pi sx}] \\ &= Ae^{-4\pi^2 s^2 t/rc}\cos\,(2\pi sx + \Phi). \end{aligned}$$

The solution to the one-dimensional diffusion problem has now been presented in two forms. In the first we follow the evolution of a point concentration of heat and express the diffusion of an arbitrary distribution by linear superposition of the Gaussian humps resulting from each element of the distribution. In the second we make a Fourier analysis of the arbitrary distribution and follow the fortunes of each harmonic component as it decays exponentially at its own rate (Fig. 18.5), then resynthesizing to obtain the result.

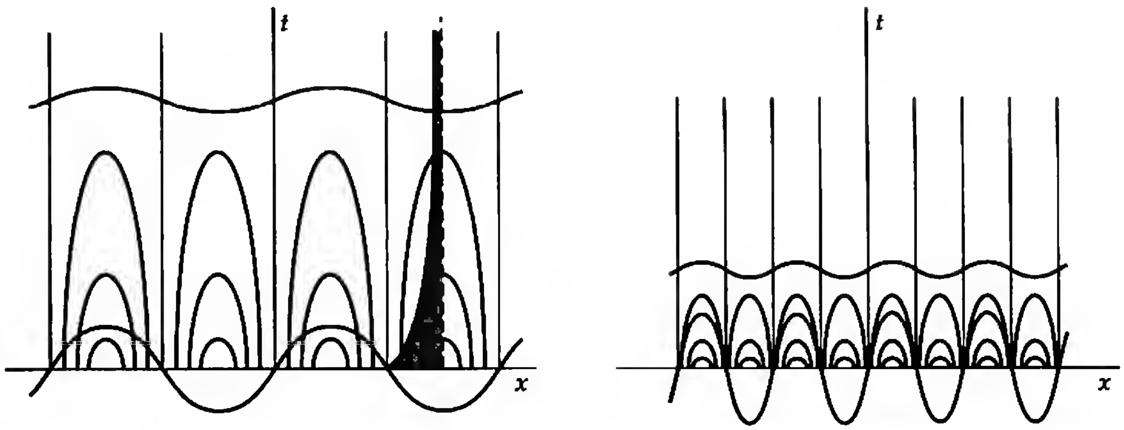


Fig. 18.5 **Exponential decay of sinusoidal distribution, without change of form, at rates proportional to the square of the spatial frequency.**

From what has gone earlier we know that the procedure in terms of convolution is one which is conceptually attractive. Both ways of looking at the matter are, however, illuminating, and we have shown that they are intimately related through the convolution theorem.

The equivalent electrical circuit suggests a number of further points. Both I and V, the two quantities necessary to specify the state of an electric circuit, are naturally regarded as equally important. However, on eliminating V from the basic equations we find that I satisfies the same differential equation as V; that is,

$$\frac{\partial^2 I}{\partial x^2} = rc\frac{\partial I}{\partial t}.$$

Thus the distribution of I behaves with the lapse of time exactly in the manner already described for V. For instance, the equation

$$I_t(x) = \Gamma_t(x)$$

describes the solution to a problem in which the flow is distributed in a Gaussian manner and was initially concentrated at one point only; in other words, the problem is what happens when V is a step function having nonzero gradient at only one point. The solution for V must be

$$-\int r\Gamma_t(x)\,dx.$$

By the principle of duality the inductive-conductive transmission line (Fig. 18.6) obeys the same equations we have been discussing, but with V and I interchanged. Thus V is proportional to the gradient of I and not vice versa.

Another possible transmission-line circuit suggests itself, namely the capacitive-conductive one shown in Fig. 18.7. Its equations will be

$$\frac{\partial}{\partial t}\left(\frac{\partial V}{\partial x}\right) = SI$$

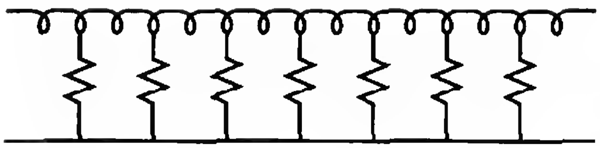


Fig. 18.6 **The inductive-conductive line.**

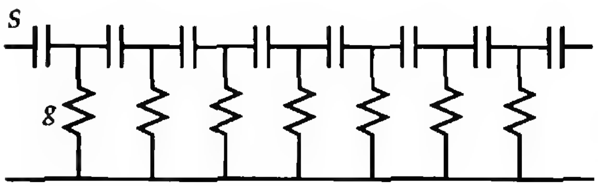


Fig. 18.7 **The capacitive-conductive line.**

and
$$\frac{\partial I}{\partial x} = gV,$$
whence
$$\frac{\partial^3 V}{\partial x^2 \partial t} = SgV,$$

and I obeys the same third-order partial differential equation. Since one might expect the simple combination of elements to behave rather simply, it is perhaps surprising to see a third-order equation appear where second-order sufficed previously. Going immediately to the initial sinusoidal distribution for a clue to the general character of the phenomena, we have

$$(i2\pi s)^2 \frac{d\mathbf{V}}{dt} = Sg\mathbf{V},$$
whence
$$\mathbf{V} = Ae^{i\Phi}e^{-Sgt/4\pi^2 s^2}.$$

Thus a uniform distribution of V would decay instantaneously, and harmonic spatial distributions would decay with time constants proportional to the square of the spatial frequency, extremely fine spatial variations being persistent. By the principle of duality we can say that the resistive-inductive line of Fig. 18.8 obeys the same differential equation, the roles of V and I being interchanged.

Solutions to the diffusion equation subject to more general initial and boundary conditions have traditionally been obtained by application of Laplace transform formalism (see Carslaw and Jaeger, 1947; van der Pol and Bremmer, 1955).

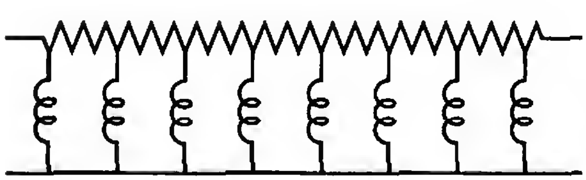


Fig. 18.8 **The resistive-inductive line.**

SINUSOIDAL TIME VARIATION

In the a-c theory of transmission lines it is known from experience that the impedance, that is, the ratio of V to I, and the propagation constant provide very convenient secondary parameters which for many purposes are more useful than the primary ones, in our case r and c. We could expect this superior utility to carry over into problems in which alternating excitation arises in heat-diffusion problems. The characteristic impedance Z_0 is given by

$$Z_0 = \left(\frac{r}{i\omega c}\right)^{\frac{1}{2}},$$

where ω is the angular frequency of excitation and the propagation constant γ is given by

$$\gamma = (i\omega cr)^{\frac{1}{2}}.$$

Both Z_0 and γ are complex, but since each has its real and imaginary parts equal in absolute value, only two real constants are involved (corresponding to the two real data r and c).

BIBLIOGRAPHY

Carslaw, H. S., and J. C. Jaeger: "Conduction of Heat in Solids," Oxford University Press, London, 1947.

Jakob, M.: "Heat Transfer," Wiley, New York, 1957.

van der Pol, B., and H. Bremmer: "Operational Calculus Based on the Two-Sided Laplace Integral," Cambridge University Press, Cambridge, UK, 1955.

PROBLEMS

1. **Retrodiction.** The temperature distribution in a rod is given. It is required to retrace its history and to discover the initial (or some earlier) temperature distribution. We know that the diffusion equation does not govern the flow of events in reverse. Derive the relevant differential equation. Consider the possibility of inverting the convolution process used for proceeding forward in time, and apply the concept of filtering spatial components to the problem. ▷

2. The temperature distribution on an infinite bar for which $rc = 2500$ seconds/square meter is given by

$$V(x) = 300 + 10e^{-300x^2} + 5e^{-75x^2}.$$

Deduce the probable past history of $V(x)$.

3. Compare and contrast the reverse diffusion problem with the problem of equalization by means of filters. ▷

4. **Fluctuation of thermal radiation.** The surface temperature at the nearest point of the moon is approximately 350 degrees Kelvin for half the month and 150 degrees Kelvin for the remainder. (Infrared measurements at times of lunar eclipse show that the surface temperature can fall at a rate of hundreds of degrees Kelvin per hour when the sunlight is turned off.) The average temperature of the moon's disk at a wavelength of one centimeter is observed to be $240 + 40 \cos(\omega t - \frac{1}{4}\pi)$ in degrees Kelvin, where $\omega = 2.46 \times 10^{-6}$ radians/second. Explain qualitatively (*a*) why the variation is cosinusoidal; (*b*) why the amplitude is only 40 degrees Kelvin; (*c*) why full moon at microwavelengths comes an eighth of a month later than full moon as optically defined. ▷

5. **Electric-thermal analogy.** The ohmic dissipation term $\frac{1}{2}rI^2$ is a moment-by-moment and point-by-point expression of the irreversibility of diffusion of charge in an electric cable. Interpret $\frac{1}{2}rI^2$ in the language of heat conduction, identifying it with some thermodynamically irreversible process. ▷

6. The increase in entropy dS when a quantity of heat dq is absorbed by unit length of a bar, where the temperature is V, is defined by

$$dS = \frac{dq}{V}.$$

Satisfy yourself that the entropy of the whole bar must increase steadily with the progress of diffusion, and prove that the increase in entropy as the temperature distribution changes from $V_0(x)$ to $V_t(x)$ is given by

$$c \int \log \frac{V_t(x)}{V_0(x)}\, dx = c \int \log V_t(x)\, dx + \text{const.}$$

Now interpret the expression $\int \log V_t(x)\, dx$ in terms of voltage distributions over electric cables, and identify it with some concept from the field of electrical communications which would be expected to increase with the passage of time. ▷

7. **Cylindrical diffusion.** A meteorite shooting through the earth's atmosphere leaves a trail of α electrons and positive ions per meter, which diffuse away with a diffusion coefficient K. Show that the electron density N per cubic meter at a distance r from a point on the meteor trail at a time t after the meteorite passes the point is given by

$$N = \frac{\alpha}{4\pi Kt} e^{-r^2/4Kt}.$$

If the refractive index of air containing N electrons per cubic meter is given, for radio waves of frequency f, by

$$\left(1 - \frac{81N}{f^2}\right)^{\frac{1}{2}},$$

show that the surface of zero refractive index is a cylinder whose radius expands to a maximum radius of

$$\frac{3\alpha^{\frac{1}{2}}}{f}$$

and shrinks to zero after a total time

$$\frac{6.2\alpha}{Kf^2}. \triangleright$$

8. A weakly ionized column contains α electrons per unit length distributed so that the number per unit area distant r from the axis is $N(r)$. Radio waves of length λ incident normally on the column are partially reflected. Show that as the column diffuses, interference reduces the echo power by a factor

$$\left|\frac{1}{\alpha}\int\int N(r)e^{i2kx}\, dS\right|^2,$$

where k is the propagation constant $2\pi/\lambda$, x is measured from the axis in the incident direction, and dS is an element of area in the transverse plane. ▷

9. **Thermal-electrical analogy.** Seek a one-dimensional continuous physical system analogous to the electrical transmission line whose equation was $\partial^3 V/\partial x^2\, dt = V$. ▷

10. **Half-order derivative.** Show that the flow of heat into a bar is proportional to the half-order derivative of the temperature difference applied at one end.

11. **Spherical diffusion.** You mix some puddings from an old recipe and hang them in a cloth to drain until they are stiff, spherical, homogeneous balls weighing one kilogram each. Day after day you pop one into a preheated oven and increase the cooking time until one day inspection reveals no soggy uncooked core. You now prepare a ten kilogram pudding. Explain why you think the proper cooking time, after the oven is preheated exactly as before, will be five times as long.

12. **Diffusion.** At each of a number of places equally spaced by a distance d on a long ditch of length L, I simultaneously (at $t = 0$) drop a quantity Q of liquid fertilizer which rapidly diffuses through the stagnant water in the ditch. Let $V_t(x)$ be the concentration of fertilizer in kilograms/meter as a function of distance x along the ditch at time t.
   (*a*) Write an expression for the initial distribution $V_0(x)$.
   (*b*) In time, the concentration will be nearly uniform. What constant value will it approach?
   (*c*) If $I_t(x)$ is the number of kilograms per second flowing in the positive x direction at a cross section x at time t, is it true that $\partial V_t(x)/\partial t = -\partial I_t(x)/\partial x$?
   (*d*) If we Fourier-analyze the initial distribution $V_0(x)$, what do we get for the initial amplitude of the Fourier component whose wavelength is d (spatial frequency $1/d$)? Check your units to make sure that the amplitude you obtain is in kilograms/meter. (You may find it advantageous to solve this problem for an infinitely long ditch, which should give the same answer.)
   (*e*) We allow a long time to elapse, so that the concentration is not quite uniform, but we would like to consider just how irregular it is. Write an approximate expression for $V_t(x)$, defining any additional symbols you have to introduce.

13. **Sinusoidal temperature variation.** The temperature $\theta(t)$ in the soil in degrees Fahrenheit, a few feet below the surface, is given by $\theta(t) = 60° + 10° \cos(\omega t - \pi/4)$, where $2\pi/\omega = 1$ year, and $t = 0$ corresponds to midsummer.

(*a*) Give reasons to believe that the vertical heat flow $I(t)$ at the same depth in response to the applied temperature variation would also depend sinusoidally on time.

(*b*) Assume that $I(t)$ does vary sinusoidally, say $I(t) = I_0 \cos(\omega t + \alpha)$. Deduce the phase angle α.

(*c*) In view of the fact that the soil temperature $\theta_0(t)$ at the surface shows pronounced diurnal variations, and thus is *not* of the simple form $A + B\cos(\omega t + \theta)$, where ω has the same value as above, can you explain why the subsurface temperature produced by the irregular surface input excitation would tend to simple annual sinusoidal form? ▷

14. **Transmission-line analogue.** Describe a mechanical one-dimensional continuous system that obeys the one-dimensional diffusion equation. Find an example where such a system might arise. ▷

15. **Subsurface temperature.** On a sunny day the temperature at the surface of the soil rises rapidly, reaches a maximum in the early afternoon, and falls sharply around sunset. Explain why the soil temperature at a depth of 1 m varies essentially sinusoidally with a period of 24 h and has a peak-to-peak variation much less than the maximum-to-minimum temperature change at the surface. As representative values, take the soil density ρ as 2000 kg m^{-3}, the specific heat s as 50 J kg^{-1} K^{-1}, and the thermal conductivity k as 30 J m^{-1} s^{-1} K^{-1}. ▷

19

Dynamic Power Spectra

One can represent an elaborate waveform, such as might be generated by a microphone responding to a passage of music, by its Fourier series, or spectrum, and it can be claimed that any subtlety in the original passage, even the briefest pizzicato, is captured in the spectrum, the spectrum being fully equivalent to the waveform. But we would have trouble pointing out a feature in the transform that corresponds to plucking the string; and if we could find such a feature, it would be hard to say, just by looking at the spectrum, at what moment the violin string was plucked. That is because the Fourier series coefficients are constants, independent of time. If we had only the power spectrum it would be impossible to say when the string was plucked because the power spectrum is independent of choice of the origin of time and does not change if some harmonics are translated in time with respect to others. It follows that the epoch information must be encoded in the phase of the complex coefficients distributed around the natural frequency of the string (196 Hz if the G-string was plucked). The phase of the 196 Hz coefficient alone would not suffice, because that phase would be compatible with a time shift of any number of natural periods.

THE CONCEPT OF DYNAMIC SPECTRUM

The ear is equipped for effectively analyzing sound into a bank of frequency bins by the use of several thousand flexible hair cells of graded sizes excited by a tapered fluid-dynamic waveguide, the cochlea. But the ear does not merely perform a Fourier analysis; the function of the operations performed behind the ear is to extract information while the music is in progress. So the ear's frequency-analysis filter bank continually furnishes changing individual outputs from which we construct an impression of the actions that originally created the sound. The ear effec-

tively attends to the signal in short successive time intervals. When high-frequency resolution is needed, as when a violinist plays a drawn out note to match the oboist's standard 440 Hz, long time segments are automatically attended to. When high time resolution is needed, as for recognition of plosive consonants, attention shifts to shorter time segments. These changes are made adaptively as the signal flows in.

Musical notation suggests very aptly the sensation of pitch and duration experienced by the listener. The notes of high pitch are high on the staff, while time flows from left to right. A dynamic spectrogram inherits these two features of musical notation.

To describe a spectrum that develops with time involves a compromise between spectral and temporal resolution that is connected with the uncertainty principle of quantum mechanics, a relationship originally emphasized by Dennis Gabor (1900–1979) (see Gabor, 1946), the father of holography. Fine spectral resolution requires segmentation into longer time slices; fine time resolution results from short segments, and thereby accepts lower frequency resolution.

Computed presentations of recordings of birdsong, speech (Rabiner and Schafer, 1978), and various geophysical phenomena have usually been based on division of the time signal into short equal segments, each segment then being Fourier analyzed into complex coefficients. On each segment of the t-axis, cells are stacked vertically to contain the coefficients (Fig. 19.1). The cell height (or bandwidth), is inversely proportional to cell width (or duration) in accordance with the uncertainty relation; thus the cell shape, or aspect ratio, is the same all over the computed presentation. However, the aspect ratio would be different if shorter segments were chosen so as not to blur brief transients or if, conversely, longer segments were chosen in order to gain better discrimination in frequency. Other ways of partitioning the (t, f)-plane will be dealt with after techniques of measurement and computation have been explained and some dynamic spectra have been illustrated.

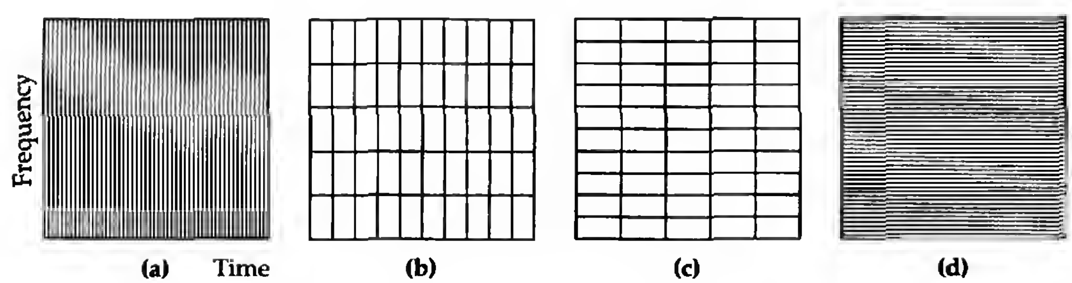


Fig. 19.1 **Partition of the (t, f)-plane into cells of width Δt and height Δf that cannot have an area less than a certain limit. Each cell is associated with a complex number that in (a) is a time sample of the signal and in (d) is a transform value. Cases (b) and (c) illustrate frameworks for dynamic spectra with different choices of cell height-to-width ratio.**

■

THE DYNAMIC SPECTROGRAPH

Fig. 19.2 shows a dynamic spectrum on the (t,f)-plane constructed from a recorded waveform that is too complex to interpret by eye. However, the ear hears the waveform as clicks (vertical streaks of short duration and wide frequency spread) and long whooshes of descending pitch. Fig. 19.3 shows a representation of speech. Seismograms, radar echoes from planetary surfaces and the Earth, electrocardiograms, and other time signals can be presented similarly. In each case visual analysis stimulates thought about the physical causes. Such dynamic spectra are obtained with instruments that work in principle as shown in Fig. 19.4. The signal waveform $s(t)$ to be analyzed is amplified, the signal is divided N ways, sequentially or simultaneously, and the same $s(t)$ is passed to a bank of N bandpass filters centered on suitably spaced frequencies f_j covering the desired range. The center frequencies are uniformly spaced and all the filters have the same bandwidth, as indicated by the filter characteristics plotted against the leftmost frequency axis. The output from each filter oscillates close to the center frequency appropriate to that filter with an amplitude and phase that vary in time in

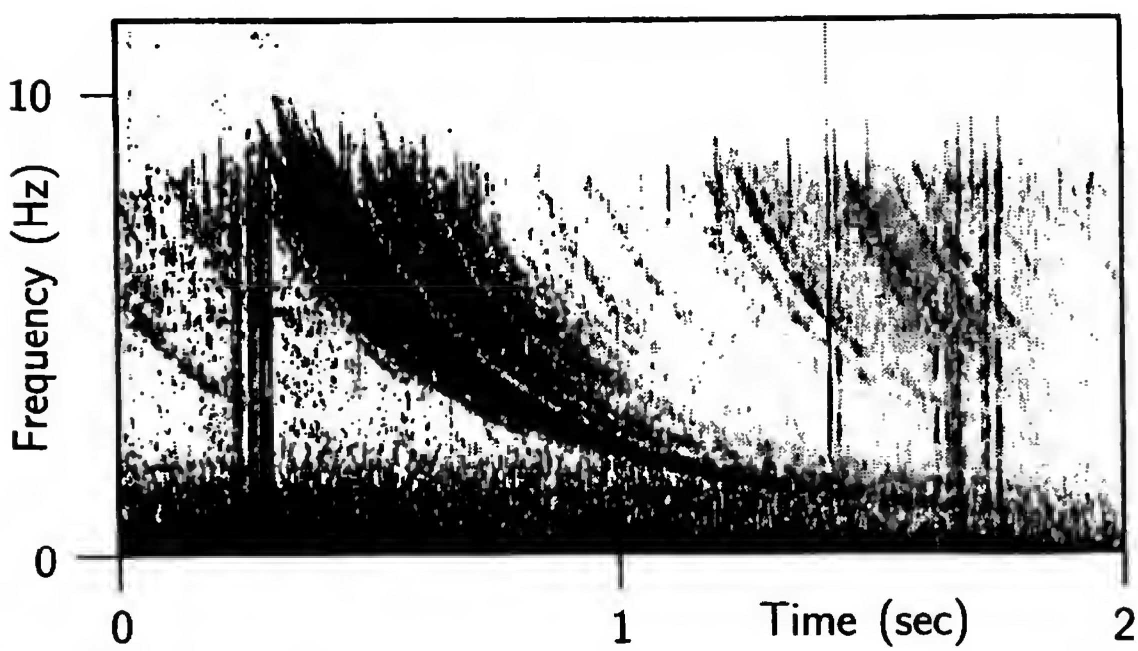


Fig. 19.2 **A dynamic spectrum of natural phenomena in the VLF band, recorded at Palmer Station in the Antarctic, showing lightning flashes (vertical traces) and whistlers (curves), which are lightning signals that have been dispersed by propagation through space many earth radii away before returning to Earth. The many physically significant features would not be apparent from the original waveform record as a function of time alone. Courtesy Jerry Yarbrough.**

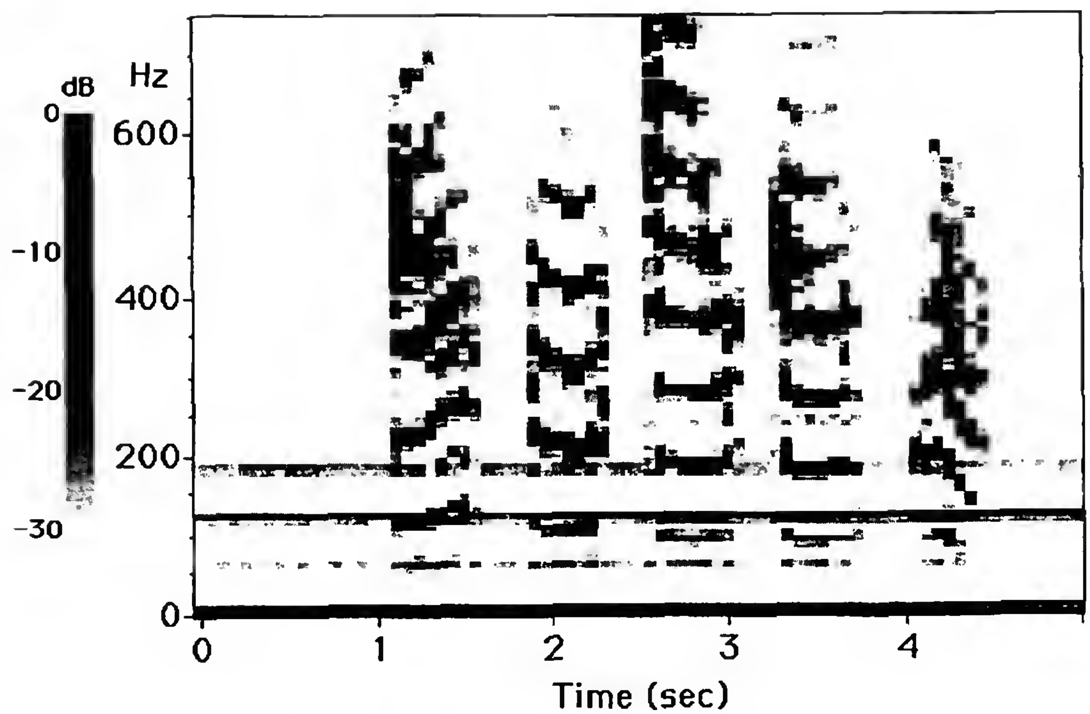


Fig. 19.3 **A voice print of "bait beat bite boat boot." The vowels are characterized by a duration of about one-third of a second and the possession of many harmonics of significant strength. The consonants are brief and broadband. Continuous hum at 60, 120, 180, and 240 Hz can be seen, and shows amplitude modulation associated with the speech. The decline in fundamental frequency (around 100 Hz) from one vowel to the next is not associated with intonation but is characteristic of the vowels themselves, and can be heard if listened for. The decline in frequency between utterances becomes more noticeable in the overtones. Frequency variation within each utterance is what determines the vowel sounds which, if listened to, would identify the accent of the speaker. Courtesy Jerry Yarbrough.**

accordance with the signal content in its band as time elapses. Each output signal is rectified and smoothed, with the results indicated on the right of the figure, which are then recorded. In practice the dynamic spectrum is recorded as a photographic density on film, or presented as a brightness distribution on a screen, or recorded digitally. Naturally there are a variety of ways of implementing these steps and several kinds of instruments are available commercially. Analogue instruments, by the 1950s, were of two kinds typified by the Kay Electric Sonograph which scanned a tunable filter from 85 Hz up to 8 kHz and the Raytheon Rayspan which used a bank of 420 filters spanning any 10 kHz band in the range up to 50 kHz. The manufacturers quoted comparable bandwidths (around 40 Hz) and resolution times (10 ms). The steps of rectification and smoothing mean that no phase information is retained.

The modern example of a dynamic spectrogram shown in Fig. 19.2 was obtained by first digitizing a signal at a 15 kHz sampling rate and then processing

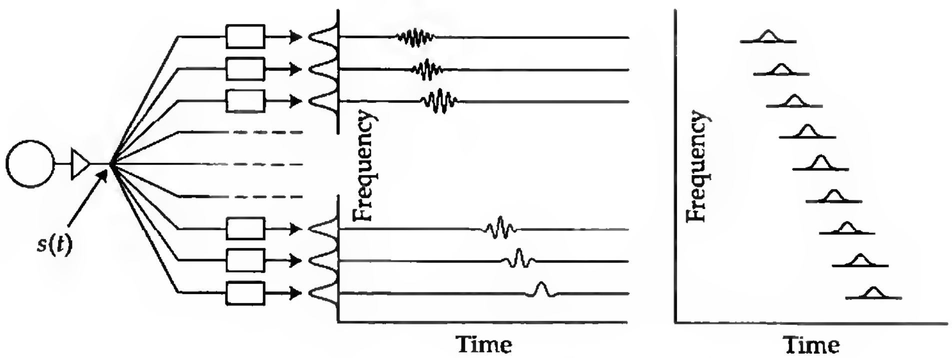


Fig. 19.4 **A signal source $s(t)$, amplifier, power divider, and the outputs $r_j(t)$ from a bank of bandpass filters centered on f_j. The filter transfer functions $H_j(f)$ shown on the left frequency axis all have the same shape; the filter impulse responses are $h_j(t)$. On the right, the dynamic spectrum is presented by profiles generated by rectifying and low-pass filtering the bandpass filter outputs. Square-law rectification will yield values of the dynamic power spectrum at points (t, f_j).**

by computer. Resolution is similar to that of the original analogue instruments. Dynamic power spectra have long been important tools in ionospheric and magnetospheric studies (Helliwell, 1965), speech analysis (Flanagan and Cherry, 1969), music (Pierce, 1983), birdsong, bat chirps, the calls of marine mammals, and other underwater sound.

Some quantitative aspects of dynamic spectral analysis are illustrated by the artificial example of Fig. 19.5. A signal waveform shown at the bottom rises and falls in strength and has a duration of 0.1 s. During the course of the signal the frequency, in some intuitive sense, falls noticeably as witnessed by the increasing peak-to-peak spacing exhibited by the waveform as time elapses. The signal was constructed from the expression $E(t)\cos[2\pi(\alpha t + \frac{1}{2}\beta t^2 + \frac{1}{3}\gamma t^3)]$, where $E(t)$ is the rising and falling envelope and $\alpha = 2200\,\text{Hz}$ is the initial frequency at $t = 0$. By differentiating the phase $\alpha t + \frac{1}{2}\beta t^2 + \frac{1}{3}\gamma t^3$ (measured in cycles) with respect to t, we get the instantaneous frequency $f_{inst} = \alpha + \beta t + \gamma t^2$. The figure uses $\beta = -27{,}500$ Hz s^{-1} for the drift rate. The term containing $\gamma = 95{,}000$ Hz s^{-2} introduces some curvature into the drift as witnessed by the descending curve of f_{inst} versus time.

The power spectrum of the signal, shown against the frequency axis, ranges from 500 Hz to 2200 Hz approximately, is somewhat skew, and as previously noted, fails to convey the drifting character of the signal. If spectral phase is available, or the Hartley transform $H(f)$ as shown on the left, inversion is possible.

Each filter has an equivalent bandwidth of 100 Hz; consequently the envelopes of the filter outputs all have the same equivalent width (0.01 s). The carrier frequencies can be seen to agree with the center frequencies read from the frequency axis. For purposes of the illustration the center frequencies are spaced 200 Hz, which is twice the bandwidth, but it would be more practical to use a spacing of one bandwidth. To use a spacing narrower than the bandwidth would be redundant, but could improve a grey-scale presentation such as Fig. 19.2.

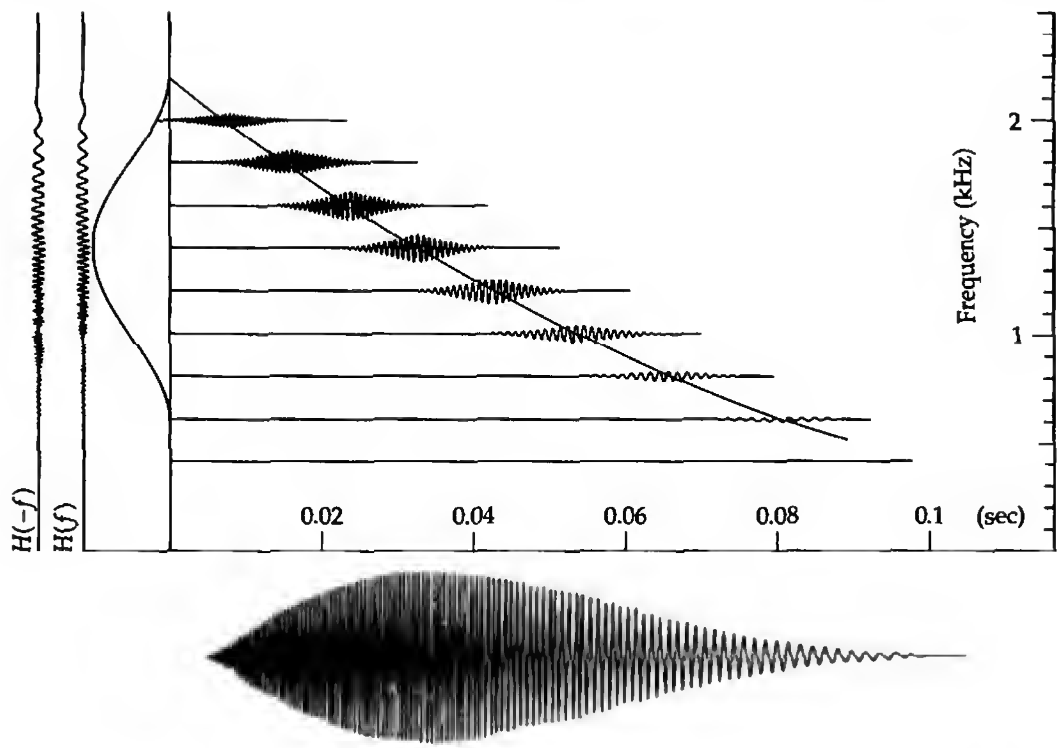


Fig. 19.5 **A signal waveform (below) and the responses of a bank of filters illustrating the frequency-division approach. The smooth curve on the left shows the power spectrum of the signal as a whole. The computed instantaneous frequency is shown by the curve threading the several filter responses.**

COMPUTING THE DYNAMIC POWER SPECTRUM

There are two ways of computing values such as those plotted on the right of Fig. 19.4: frequency division and time division. We deal with frequency division first, as exemplified by the figure.

Frequency division. Frequency division means that the incoming signal is simultaneously (or sequentially) passed to a series of bandpass filters shown as boxes in Fig. 19.4. The center frequencies cover a range from f_{min} to f_{max} and may be spaced uniformly as illustrated, or uniformly in log f, which makes sense biologically.

Let the jth filter be centered on frequency f_j. The filter bandwidths would normally be equal to the frequency spacing or greater or less according to circumstances. With uniformly spaced frequencies the filter bandwidth is thus independent of frequency; each such constant-bandwidth filter will respond to an

incoming impulse by oscillating in the vicinity of the mid-frequency of its band and be modulated by an envelope which is of the same duration for all. With frequencies uniformly spaced in $\log f$ rather than f, the bandwidths, as measured in hertz, would increase in proportion to f_j; the impulse response of each corresponding constant-Q filter would oscillate near the mid-frequency f_j but with a shorter duration for higher frequencies. The number of oscillation periods under the envelope would be the same for each filter, in fact the impulse responses would be identical, but compressed in time by a factor proportional to the mid-frequency. In Fig. 19.4, if the input signal was an impulse instead of a gliding tone, the various responses would be lined up below one another because all filters would be stimulated at the same time. Evidently the input signal illustrated is a whistle of descending pitch.

An analogue filter, characterized by its impulse response $h(t)$, will provide a response $r(t) = s(t) * h(t)$. To compute the response we work with signal samples s_i taken at a sampling rate f_{samp} from the physical input signal $s(t)$. Then $s_i = s(i/f_{samp})$, where $i = 1, 2, 3, \ldots$. Represent the filter by discrete values h_i of $h(t)$ spaced at the same sampling interval $1/f_{samp}$. Then the discrete values of the output response $\{r_i\}$ of that filter are given by $\{r_i\} = \Sigma_k s_{i-k} h_k = \{s_i\} * \{h_i\}$. The summation extends over the number of coefficients, possibly 20 or 30, deemed necessary to describe the filter.

This seems straightforward but some practical concerns will be recognized. Suppose that $f_{samp} = 15$ kHz. Then frequencies as high as 7.5 kHz will be adequately sampled in theory, but in the presence of a little noise, two samples per period is not satisfactory. Consequently, perhaps 4 kHz is as high as one might expect reliable results and one would prudently test the performance on actual physical signals. But now suppose that $s(t)$ contained components at frequencies in the range 8 to 10 kHz, as might arise from those harmonics, equally spaced in frequency, that are essential to voice recognition and musical appreciation, or from unwanted noise or interference. Samples taken at 15 kHz would be indistinguishable from samples of signal components in the range 7 to 5 kHz (aliasing). Consequently, assurance would be needed that the analogue sensor with which the digital data were acquired rejected inputs with components whose frequencies exceeded half the sampling frequency.

Convolution by summing is fast computationally when one of the factors, in this case $\{h_i\}$, has relatively few terms, say not greater than about 30. But a practical concern arises from the choice of sampling interval. If the impulse response $\{h_i\}$ varies from filter to filter, then for the lower values of f_j the number of coefficients spaced at $1/f_{samp}$ is more than is needed for adequate specification of the impulse response of those filters, by one or two orders of magnitude. The cost of storing unneeded data needs to be avoided. Constant-Q filters do not present this concern.

An alternative method for convolution is to take the Fourier or Hartley transform of the signal, adequately padded by zeros, multiply by the transfer function of each filter in turn, and invert the transform. Exactly the same result will be ar-

rived at as for time domain convolution; the choice of method will depend on circumstances, perhaps on speed of computation or perhaps on computer capacity. With signals sampled at 15 kHz, which as we have seen may be good only for low-fidelity situations where only frequencies up to about 6 kHz are present, a 5-second recording means 75,000 samples, which makes convolution by transforms impractical in some contexts. Time division is thus introduced.

Time division. Instead of dividing the signal into numerous frequency bands we can begin by gating out a segment of the time signal; two such segments are shown in Fig. 19.6 at mid-time epochs $t_i = 0.03$ s and 0.07 s. Each has a duration of 0.01 s, as measured by the equivalent width of the profile, or time window, $h(t)$. One could gate out a sharp-edged slice, just as one would have taken a sharp-edged filter pass band in Fig. 19.4, but Gaussian shading is better. Thus the earlier of the two segments is given by

$$f(t) = s(t)e^{-\pi[(t-0.03)/W]^2},$$

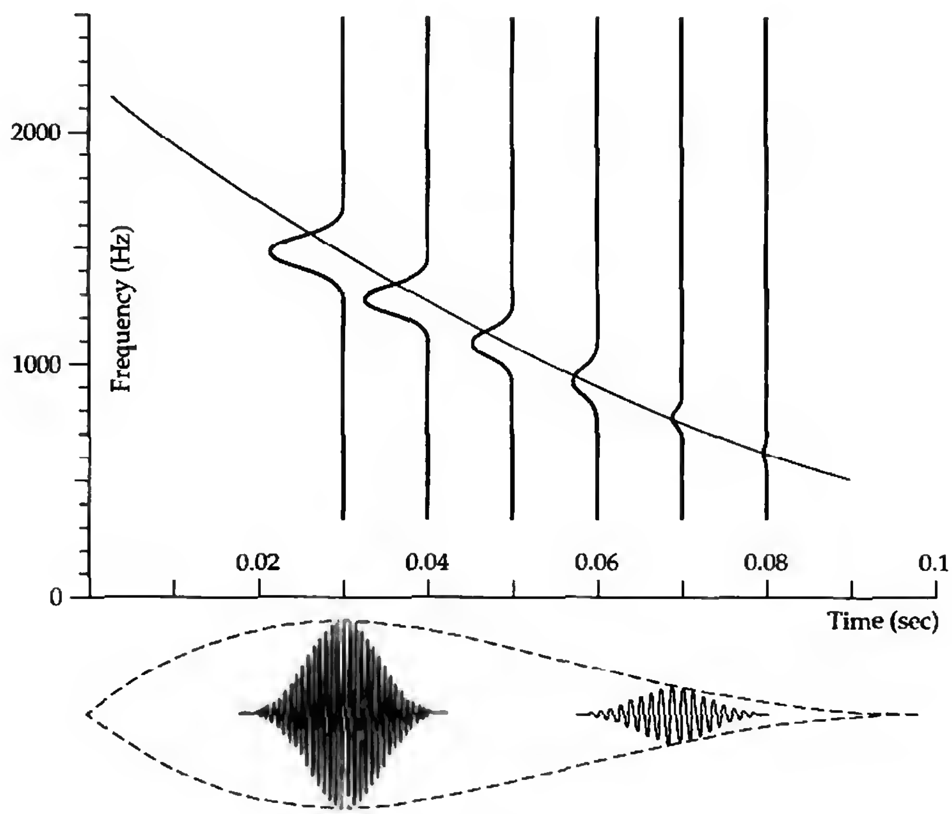


Fig. 19.6 **To illustrate time division, two segments taken from the earlier signal (broken outline) by multiplication with Gaussian functions $\exp[-\pi(t-\tau)^2/W^2]$ with equivalent width $W = 0.01$ s and $\tau = 0.03$ and 0.07 s. Power spectra of six segments are plotted against frequency (heavy curves).**

where $W = 0.01$ s. The transform of this segment can be computed, then the segment can be advanced through $W/2$ and a new transform taken, and so on until the complete signal has been covered. This is computationally convenient because all the transforms have the same length.

As an example, if $f_{samp} = 15$ kHz, then 350 samples will suffice to specify $f(t)$ over a span $2W$. At the edges of this bounded segment, the Gaussian factor is 27 dB below its peak value, which should be enough to do justice to the signal-to-noise ratio in many situations. In the figure a span of 0.02 s is illustrated, and the truncation involved can just be seen by eye. If one extended the span to $2.4W$, the envelope would be 40 dB down and there would be 420 samples. This is clearly practical, leaving leeway for broader segments and larger transforms where finer resolution in frequency is appropriate.

To show the Fourier transform of a time segment in Fig. 19.6 by the power spectrum alone means that phase has been abandoned. A dynamic phase spectrum also exists and in some applications is valuable.

Presentation. After the dynamic power spectrum values have been computed and stored there is a choice of presentations, including grey-scale renderings using pixels with different visual densities, diagrams with contours of equal value, and numerical matrix printouts. With grey-scale spectra one can enhance desired features by choosing a suitable functional relationship between the density and the computed value, set thresholds, expand the dynamic range, add false color, and so on.

EQUIVALENCE THEOREM

Let the Fourier transform in amplitude and phase be calculated at some frequency f_j for some epoch t_i by time division, and call it $G(t_i, f_j)$. Will the amplitude and phase be the same as obtained at that time t_i by frequency division with a filter centered at f_j? Clearly the answer depends on the filter bandwidth. The filter bandwidth should be such that the envelope of the impulse response corresponds to the multiplying factor used for gating out the segment centered at t_i. Narrowing the bandwidth would spread the response over a longer time and reduce the power at $t = t_i$.

In terms of frequency division let the response to the signal $s(t)$ from the filter at f_j be $r_j(t) = \int_{-\infty}^{\infty} s(t')h_j(t - t')dt'$, where $h_j(t)$ is the impulse response of the jth filter; the transfer function $H_j(f)$ of that filter is the Fourier transform of the impulse response as expressed by $H_j(f) = \int_{-\infty}^{\infty} h_j(t)\exp(-i2\pi ft)dt$. In the time division approach, one multiples the signal $s(t)$ by a translated time function $w(t - t_i)$ centered on $t = t_i$, takes the Fourier transform, and then asks whether the transform value at $f = f_j$, for the same index pair (i, j), relates to the amplitude and phase of $r_j(\tau)$. The indicated value of the Fourier transform is $\int_{-\infty}^{\infty} s(t)w(t - t_i)\exp(-i2\pi f_j t)dt$, so the two representations are equivalent if

$h_j(t_i - t) = w(t - t_i)\exp(-i2\pi f_j t)$. Thus

$$r_j(t) = e^{-i2\pi ft}\int_{-\infty}^{\infty} S(f')H_j(f' - f)e^{i2\pi f't}\,df' = \int_{-\infty}^{\infty} s(t')w(t' - t)e^{-i2\pi ft'}\,dt'.$$

Since $H_j(f)$ is of the form $H_0(f) * \delta(f - f_j)$, $f > 0$, it follows that $h_j(t)$, the inverse Fourier transform of $H_j(f)$, is indeed of the form $w(t)\exp(-i2\pi f_j t)$, and that equality results when the multiplying factor $w(t)$ is chosen to agree with the envelope of the filter impulse response.

If the pass band $H_j(f)$ is Gaussian as described by $\exp[\pi(f - f_j)^2/W_f^2]$, then the values for the dynamic power spectrum will be the same as results from time division when $w(t)$ is also Gaussian and has an equivalent width W that is the reciprocal of W_f. With $W = 0.01$ s the corresponding filter bandwidth W_f would be 100 Hz, as in fact was used in Fig. 19.5.

ENVELOPE AND PHASE

After the filter responses $r_j(t)$ have been determined the envelope (and the phase if wanted) must be found as described in connection with Fig. 19.4. Rectifying with a diode and smoothing with a low-pass filter can readily be simulated in the computer. A rough procedure for the power spectrum is to square the response, shift half a period, and add; this estimate is $[r_j(t)]^2 + [r_j(t + 1/4f_j)]^2$.

More sophisticated is to determine the instantaneous envelope and phase by taking the Hilbert transform of the response (Chapter 13). Although Fig. 19.4 shows a bank of j discrete filters, and the description of Fig. 19.6 also assumes that j is discrete, one is in fact free to think of j as a continuous variable and indeed the original Sonograph scanned continuously in frequency. Thus $r_j(t)$ can be defined in terms of the impulse response $h_j(t)$ of the jth filter as

$$r_j(t) = h_j(t) * s(t) = [e^{-\pi(t/W)^2}\cos 2\pi f_j t] * s(t),$$

where t and f_j are regarded as continuously varying and the impulse sequence $\{h_j\}$ is replaced by the impulse response $h_j(t)$. For any t and f_j, $r_j(t)$ has an instantaneous envelope $G(t, f_j)$ and phase $\phi(t, f_j)$ such that $r_j(t) = \mathrm{Re}\{G(t, f_j)e^{i[\phi(t, f_j) + \omega t]}\}$. We could call $Ge^{i\phi}$ the Gabor dynamic spectrum because, evaluated at discrete values of t and f_j, it is the same as the complex coefficient associated with a cell. In this notation the dynamic power spectrum would be $[G(t, f_j)]^2$.

Though much information may be encoded in the phase, phase is not as easy to present or interpret as the power spectrum. One way is to construct a half-tone display directly from $[r_j(t)]^2$ without smoothing. The diagram will then acquire hologram-like fringes in the form of thin, quasiparallel isophase lines passing through the zero crossings of $r_j(t)$.

Let r_j be evaluated only for discrete values of t equal to $m\Delta t$, where Δt is the width of a Gabor cell, let the values of j be related to frequency by $f = nj\Delta f$, where Δf is the critical cell height corresponding to Δt, and let the complex Ga-

bor coefficients for the cells be written a_{mn}. Gabor had in mind that the signal $s(t)$ would be expressible in terms of these coefficients by

$$s(t) = \sum_m \sum_n a_{mn} e^{-(t-m\Delta t)^2/W^2} e^{j2\pi nt\Delta f};$$

for a rigorous result, m and n need to be fractions less than unity (Bastiaans, 1994).

Much literature has descended from the paper of Gabor (1946); refer to Boashash (1988) and Cohen (1989) for bibliographies and to the Science Citation Index under Gabor and later authors.

USING log f INSTEAD OF f

We now return to nonuniform partitioning of the (t,f)-plane, first considering the example of log f instead of f on the vertical axis, exactly the same coordinate system as is used on a player-piano roll and, before that, on the cylinder of a music box. To generate a dynamic power spectrum of this kind, on a frequency-division basis, calls for a bank of filters of constant Q with impulse responses such as

$$h_j(t) = e^{-\pi t^2/T^2} \cos 2\pi f_j t,$$

where the equivalent width T is inversely proportional to the channel center frequency f_j. Thus $T = k/f_j$. Choose $k = 2$; then there are exactly two cycles at the frequency f_j in a time T. A set of such impulse responses at frequencies $f_j = 500, 707, 1000, 1414, 2000$ Hz is shown on the left of Fig. 19.7. Since the duration of each impulse response is now inversely proportional to frequency f_j, the time resolution has improved as the frequency resolution has deteriorated. When the same signal as that of Fig. 19.5 is presented to these filters the responses $r_j(t)$ are as shown.

On the 500 Hz channel the impulse response $h_{500}(t)$ is of relatively long duration, perhaps 10 ms, while the response $r_{500}(t)$ runs on for about 30 ms. On the 2000 Hz channel the impulse response is 4 times briefer, corresponding to the enhanced time resolution; however, the response $r_{2000}(t)$ is no briefer than at 500 Hz, running on for about 40 ms because of the degraded frequency resolution that accompanies the enhanced time resolution. Thus, response is spread over the time it takes for the signal tone to slide through the relatively broad frequency response $H_{2000}(f)$ of the highest frequency filter.

It is noticeable that the peak response of $r_{2000}(t)$ occurs later than the moment when the instantaneous frequency of the signal passes through the 2000 Hz center frequency of the filter. Conversely, if one focuses on a time, the frequency that reaches maximum response at that time is less than the instantaneous frequency for that time. These significant effects are a direct consequence of the uncertainty inherent in the bandwidth-duration inequality. To understand this in terms of an analogy, imagine a skewnormal bivariate probability distribution. There is a *ridge* line coinciding with the major axes of the concentric elliptical loci of constant probability density $p(x, y)$, which is analogous to our locus of instantaneous fre-

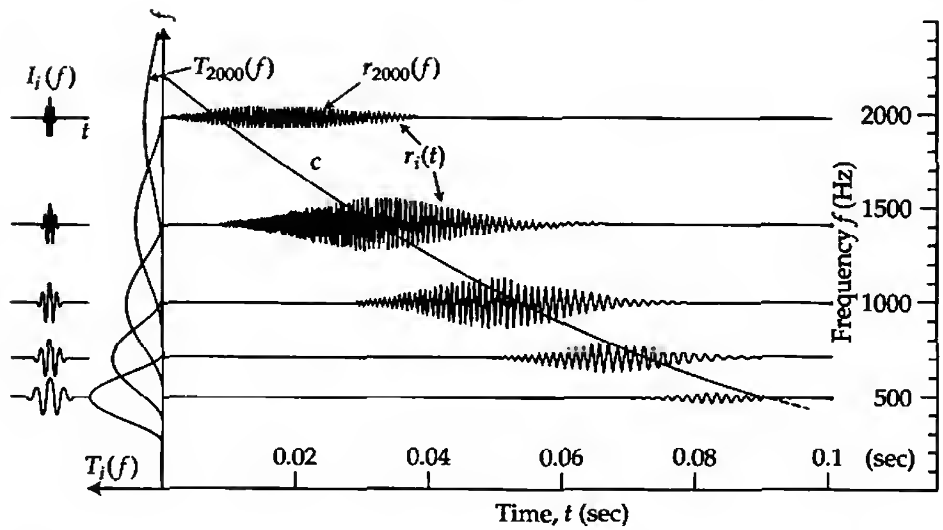


Fig. 19.7 **The responses of a bank of constant-Q filters whose center frequencies f_j, and associated bandwidths, increase in geometric progression with a frequency ratio $\sqrt{2}$. Left: Impulse responses $h_j(t)$ of the filters and their transfer functions $H_j(f)$. Main plot: The responses $r_j(t)$ to a gliding tone whose frequency descends as indicated by curve C.**

quency. The *regression* lines of y on x and of x on y are analogous to peak time versus frequency and peak frequency versus time.

In some applications, such as chirp radar, an attempt is made to determine instantaneous frequency and frequency drift rate to better precision than the bandwidth-duration inequality would seem to allow, for example by detecting zero crossings. This can be done successfully with electronically generated chirps, but if the character of the signal were not known in advance, as in observational geophysical research, then the inherent uncertainty would reveal itself when the zero-crossing technique was attempted in the presence of signals more complicated than one chirp.

■

THE WAVELET TRANSFORM

A substantial literature (Grossman and Morlet, 1984; Kronland-Martinet, 1987; Mallat, 1989; Strang, 1989; Daubechies, 1991; Sheng, 1996) that deals with dynamic spectrum analysis depending on constant-Q filters is characterized by the use of the term wavelet, which has been given a sharp meaning as follows. Instead of the set of responses

$$r_j(t) = \int_{-\infty}^{\infty} h_j(t - t')s(t')dt' = h_j(t) * s_j(t)$$

that we generate by convolving various impulse responses with the signal $s(t)$, consider the different set generated by

$$h\left(\frac{t}{\log 2\pi f_j}\right) * s(t).$$

In this reformulation, the impulse responses change with f_j but simply by time dilation of a chosen basic wavelet $h(t)$, which might be a Gaussian wavepacket but need not. The basic wavelet is chosen to be compact in time and in frequency, two conditions that are met by arranging that $h(t)$ is a square-integrable function (thus dying out in both directions) and has zero mean (frequency content not reaching to d.c.). The impulse responses generated by dilation are called wavelets.

Using filter banks for the purpose of separately encoding signal bands naturally suggested sampling rates that varied according to the band center and gave rise to "subband coding" (see Vetterli, 1984), a technique with evolutionary roots in work of Kretzmer (1946). "Wavelet" terminology has now supplanted earlier expressions of the same concepts.

In the customary notation, the wavelet transform is written

$$\Phi(a, b) = \frac{1}{|a|^{1/2}} \int_{-\infty}^{\infty} h\left(\frac{t-b}{a}\right) s(t) dt.$$

The quantity b having the dimensions of time represents translation in time, while the scale factor a, defined as $-\log 2\pi f_j$, also having dimensions of time, represents dilation, or stretching, in time. The normalization factor $|a|^{-1/2}$ compensates for the stretching to give the wavelets all the same quadratic content. As frequency increases, a becomes more and more negative, so the corresponding frequency axis points downward instead of upward. The definition of $\Phi(a, b)$ contains a cross-correlation rather than a convolution integral, a distinction that would not matter if $h(\,)$ was an even function. Nevertheless, the wavelet transform can be introduced as a convolution, as was done above, by dealing directly in terms of f_j instead of a.

The wavelet transform is thus a highly oscillatory entity and if rectified and directly displayed as a half-tone image would present the striations discussed above in connection with phase display. Of course, the wavelet transform may be smoothed after rectification and presented in the same manner as in Fig. 19.2. The main difference would lie in the broadening of the filter bandwidths as $|a|$ increased. If $h(t)$ was chosen as a Gaussian wavepacket, as in Fig. 19.7, that would be the only difference. As explained above, if one chose to plot on the (a, b)-plane, the diagram would be inverted and would also be compressed nonlinearly in the vertical direction so that a linear frequency drift would appear curved.

Although the wavelet terminology in its origins simply connoted constant-Q filters as distinct from the constant-bandwidth filters discussed by Gabor, several mathematical advances of wide-reaching significance quickly occurred. For one thing, as Gabor had pointed out, Gaussian elementary signals are not orthogonal. This deficiency has been overcome by making use of the freedom of choice of the basic wavelet.

Ability to invert the wavelet transform has become of importance in data compression, especially of images. When a waveform is represented by cells on the (t, f)-plane, the omission, before transmission, of cells with little content, or threshold coding, can clearly compress the data to some extent. The practice in subband coding was to omit components beyond a certain level of decomposition, a practice that has been refined under wavelet theory and competes with the DCT encoding of 8×8 blocks associated with the now superseded JPEG4 image transmission protocol. Thus the subject of dynamic spectra has broadened beyond the simple representation of complicated waveforms by visually interpretable patterns on the (t, f)-plane. Nevertheless, simple representation based on Gaussian elementary signals will continue to be widely practiced, especially as extended by the use of spatially variable cell shapes, chirplets, and algorithms that can adapt locally to the signal encountered and bring into focus fine structure that has hitherto been blurred.

■

ADAPTIVE CELL PLACEMENT

Constant-Q filters have been advocated for the purpose of obtaining an increase in time resolution at high frequencies, and in the case of whistlers of descending pitch and diminishing frequency drift a better representation would result from improved frequency resolution at low frequencies combined with improved time resolution at higher frequencies. In the two schemes described so far the cell shape and placement did not take the character of the signal into account, the cell shape not allowing for local structure on the (t, f)-plane. But there is no reason why the cell shape should not be allowed to depend both on time and frequency. One might base a specification of cell shape versus cell location on a provisional dynamic spectrogram. Of course, tiling the plane with cells of changing shape is not necessarily possible without overlap or gaps; however, there is no harm in allowing cells to overlap, as when more filters than necessary are used in order to improve presentation, although no information is added.

■

ELEMENTARY CHIRP SIGNALS (CHIRPLETS)

We have seen that the elementary cells on the (t, f)-plane may have any width and height consistent with a lower limit to area, and that the aspect ratio may be allowed to vary with frequency. When the lower limit is achieved or approximated, an elementary signal, as defined by Gabor, exists. It is just the impulse response of the analyzing filter. The cell aspect ratio may vary with time if the signal to be analyzed so suggests. That introduces filters that do not observe time invariance, and may adapt to the arriving signal as the ear does.

To take advantage of this flexibility in computing one would first analyze with a suitable choice of fixed time and frequency resolution. Where vertically striated

structure was found, analysis could be repeated with better time resolution, at a sacrifice of frequency resolution in the direction where little frequency dependence was found. Conversely, areas of horizontal striation could be redone with better frequency resolution, while oblique structure would call for chirplets.

Figure 19.2 showed the dynamic spectrum of a natural phenomenon with a multiplicity of descending tones of different strengths but of almost identical shape separated by short time delays, new components being received before earlier ones died away. At the low frequencies neighboring components appear to merge; how does one know that at higher frequencies the structures presented are not already composite but unresolved? Fig. 19.8 considers two simple gliding tones with a time separation of one second. A square cell straddling the pair of instantaneous frequency loci shows a choice of Δf and Δt that would not resolve the components. Two other rectangular cells representing other choices of elementary signals, but of the same area, show that making the filter bandwidth neither narrower (lower right rectangle) nor broader (upper left) can resolve the dual structure. However, if the (t, f)-plane could be partitioned into oblique cells as shown centered on the earlier of the two components then the structure would be resolved. To do this requires filters with impulse responses of the form

$$T^{-1}e^{-\pi(t-T)^2} {\textstyle\frac{\cos}{\sin}}[2\pi(ft + \tfrac{1}{2}\beta t^2)],$$

where the frequency drift rate β (negative in the example) could be chosen in accordance with the degree of obliquity desired. This new elementary signal, not known to Gabor, is called a chirplet and was reported independently at about the same time by Mihovilovic and Bracewell (1991, 1992) and Mann and Haykin (1992); a chirplet occupies a minimum area on the (t, f)-plane. Computed dynamic spectra for the gliding-tone pair confirm that the fine structure is resolvable using chirplets. The results also show that in any given time segment when signals of two frequencies are both present in a rectangular cell, there will be beats modu-

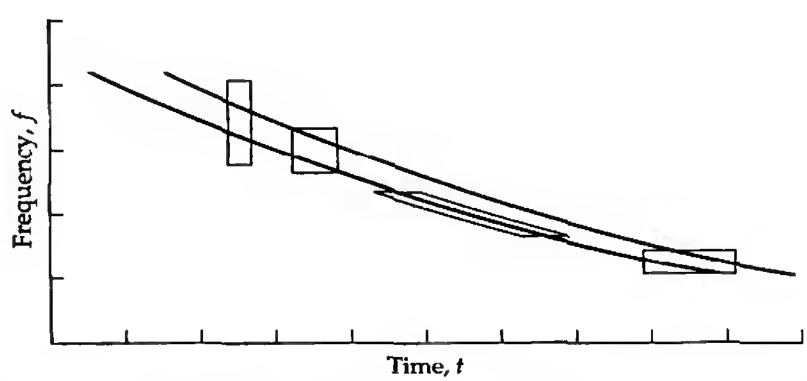


Fig. 19.8 **Two gliding tones in quick succession that would not be resolved by square cells nor by cells elongated vertically (left) or horizontally (right), as evidenced by the fact that the cells straddle both frequency loci. Oblique cells of the same area could give resolution in the direction normal to the loci.**

lating the strength of the time response of the filter. In Fig. 19.2 such beats are seen in several places, indicating that several of the apparently simple components are in fact composite. Adaptive chirplet analysis can thus be expected to be of importance in physical investigations, wherever complicated waveforms are recorded.

THE WIGNER DISTRIBUTION

Among the various ways of operating on a one-dimensional time function $f(t)$ so as to set up a two-dimensional representation on the time-frequency plane, Wigner's procedure is one of the oldest. Let the real-valued Wigner distribution $W(t, f)$ corresponding to $f(t)$ be

$$W(t, f) = \int_{-\infty}^{\infty} f(t + \tfrac{1}{2}\tau) f^*(t - \tfrac{1}{2}\tau) e^{-i2\pi f \tau} d\tau.$$

Some examples can be easily established. Thus $\delta(t - T)$, a unit impulse at $t = T$ has a Wigner distribution that is a vertical line crossing the (t, f)-plane at $t = T$. This can itself be written as $\delta(t - T)$, a distribution now to be regarded as a function of two variables, t and f, but independent of f. The Gabor diagram for a click would look much the same on a grey-scale rendering, except that the Wigner distribution is completely free of the fuzziness, associated with the uncertainty principle, that is characteristic of Gabor diagrams and the various wavelet transforms.

As a second example take a chirp signal of constant amplitude whose frequency is f_0 at $t = 0$ and increases with time at a rate β Hz s^{-1}, so that the instantaneous frequency is $f_0 + \beta t$. The Wigner distribution for $f(t) = \exp[i2\pi(\frac{1}{2}\beta t^2 + f_0 t - c)]$ is $\delta(\beta t + f_0 - f)$, a distribution concentrated on the straight line $f = f_0 + \beta t$. This is in striking agreement with the mental picture of what might happen when one steadily turns the frequency knob of a signal generator. The physical reality of the limits to resolution set by the uncertainty principle, however, would not permit an equally clear Wigner distribution for a pair of gliding tones $f(t) + f(t - \Delta t)$ nor for a single gliding tone with a Gaussian amplitude envelope.

As a third example, $f(t) = \cos 2\pi f_0 t$, a cosine wave of frequency f_0, which would appear as a horizontal band at $f = f_0$ on a Gabor diagram, has a Wigner distribution $\frac{1}{4}\delta(f + f_0) + \frac{1}{4}\delta(f - f_0) + \frac{1}{2}\delta(f) \cos 4\pi f_0 t$. In addition to the horizontal lines at $f = \pm f_0$, an interaction term has appeared, in the form of a straight line at $f = 0$, varying in strength with a period $\frac{1}{2}f_0$ as time elapses. This interaction term is a mathematical consequence of the definition and does not have the intuitive appeal of the steady lines at $f = \pm f_0$.

Finally consider $f(t) = \delta(t - 1) + \delta(t - 5)$. The Wigner distribution is $\delta(t - 1) + \delta(t - 5) + 2\delta(t - 3) \cos 8\pi f$. The horizontal lines that each impulse, taken alone, would give rise to, are there, but in addition another has appeared at the mean time of occurrence $t = 3$ and it is modulated in the f-direction.

The Wigner distribution for a complicated signal made up of many parts, such as a speech waveform, would be quite unsuitable for the visual interpretation that

the Gabor diagram is useful for.

The Wigner distribution can readily be inverted; thus, given $W(t, f)$, one obtains the original signal as

$$f(t) = \frac{1}{f^*(0)} \int_{-\infty}^{\infty} W(\tfrac{1}{2}t, f) e^{i2\pi ft} df.$$

Furthermore, projecting downward onto the time axis yields

$$|f(t)|^2 = \int_{-\infty}^{\infty} W(t, f) df$$

while projecting horizontally onto the frequency axis yields the power spectrum

$$|F(f)|^2 = \int_{-\infty}^{\infty} W(t, f) dt.$$

The total quadratic content of $f(t)$ equals the volume under the Wigner distribution:

$$\int_{-\infty}^{\infty} |f(t)|^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} W(t, f) dt df.$$

The Wigner distribution proves to have applications to optics that have distinguished it as a tool for thinking about existing systems. Consider that the Wigner distribution of the Fourier transform $F(\ )$ of the signal $f(t)$ is $W(-f, t)$, which is the same as the distribution $W(t, f)$ rotated through 90°. In a graded index fiber (Chapter 13), the spatial field distribution in the input aperture converts itself to its Fourier transform in a distance L that is calculable from the refractive index distribution in the transverse plane. At intermediate locations one ath of the distance L from the input, the field is expressible as a fractional Fourier transform of order a of the aperture distribution. Lohmann (1993) has described a direct connection with rotation of a Wigner distribution by one ath of 90°, while further connections with wavelets, chirp transforms, and optical diffraction are dealt with by Ozaktas et al. (1993).

BIBLIOGRAPHY

Bastiaans, M. J.: "Oversampling in Gabor's Expansion by an Integer Factor," Proc. IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis, Philadelphia, Pennsylvania, USA, October 25–28, 1994.

Boashash, B.: "Time-Frequency Signal Analysis and Synthesis," SPIE vol. 975, pp. 164–185, 1989.

Cohen, L.: "Time-frequency Distributions—A Review," *Proc. IEEE*, vol. 77, pp. 941–981, 1989.

Daubechies, I.: "The Wavelet Transform: A Method for Time-Frequency Localization," pp. 366–417 in "Advances in Spectrum Analysis and Array Processing," vol. 1, S. Haykin, ed., 1991.

Flanagan, J. L., and L. Cherry: "Excitation of Vocal-Tract Synthesizers," *J. Acoust. Soc. Amer.*, vol. 45, pp. 764–769, 1969.

Gabor, D.: "Theory of Communication," *J.I.E.E.*, vol. 93, Part III, pp. 429–444, 1946.

Grossman, A., and J. Morlet: "Decomposition of Hardy Functions into Square-Integrable Wavelets of Constant Shape," *SIAM J. Math. Anal.*, vol. 15, pp. 723–736, 1984.

Helliwell, R. A.: "Whistlers and Related Ionospheric Phenomena," Stanford University Press, 1965.

Kretzmer, E. R.: "Reduced-alphabet Representation of Television Signals," *IRE Conv. Rec.*, vol. 4, pp. 140–147, 1956.

Kronland-Martinet, R., J. Morlet, and A. Grossman: "Analysis of Sound Patterns Through Wavelet Transforms," *Intl. J. Pattern Recognition and Artificial Intelligence*, vol. 1, pp. 97–126, 1987.

Lohmann, A. W.: "Image Rotation, Wigner Rotation, and the Fractional Fourier Transform," *J. Opt. Soc. Am. A*, vol. 10, pp. 2181–2186, 1993.

Mallat, S.: "A Theory for Multi-resolution Signal Decomposition: The Wavelet Representation," *IEEE Trans. Patt. Anal. Machine Intell.*, vol. 7, pp. 674–693, 1989.

Mann, S., and S. Haykin: "'Chirplets' and 'Warblets': Novel Time-frequency Methods," *Electronics Letters*, vol. 28, no. 2, pp. 114–116, 1992.

Mihovilovic, D., and R. N. Bracewell: "Chirplet Analysis of Dynamic Spectra," *Electronics Letters*, vol. 27, pp. 1159–1161, 1991.

Mihovilovic, D., and R. N. Bracewell: "Whistler Analysis in the Time-Frequency Plane Using Chirplets," *Journal of Geophysical Research*, vol. 97, pp. 17199–17204, 1992.

Ozaktas, H. M., B. Barshan, D. Mendlovic, and L. Onural: "Convolution, Filtering, and Multiplexing in Fractional Fourier Domains and Their Relation to Chirp and Wavelet Transforms," *J. Opt. Soc. Am. A*, vol. 11, pp. 547–559, 1994.

Pierce, J. R.: "The Science of Musical Sound," Scientific American Library, New York, 1983.

Qian, S., and D. Chen: "Joint Time-frequency Analysis: Methods and Applications," Prentice-Hall, Upper Saddle River, New Jersey, 1996.

Rabiner, L. R., and R. W. Schafer: "Digital Processing of Speech Signals," Prentice-Hall, Upper Saddle River, NJ, 1978.

Sheng, Y.: "Wavelet Transform," pp. 747–827 in "The Transforms and Applications Handbook," A. D. Poularikas, ed., CRC Press/IEEE Press, Boca Raton, Florida, 1996.

Strang, G.: "Wavelets and Dilation Equations: A Brief Introduction," *SIAM Review*, vol. 31, pp. 614–627, 1989.

Vetterli, M.: "Multi-dimensional Subband Coding," *Signal Processing*, vol. 6, pp. 97–112, 1984.

PROBLEMS

1. **Wavelet definition.** A basic wavelet $h(t)$ has a Fourier transform $H(f)$. (a) Show that the Fourier transform of $|a|^{-1/2}h[(t-b)/a]$ has the same quadratic content as $H(f)$. (b) Alternatively, if wavelets were constructed from the basic wavelet as $|a|^{-1}h[(t-b)/a]$, show that the Fourier transforms would all have the same value at $f=0$.

2. **Scaling function.** A function $\phi(x)$ is to satisfy a *dilation equation* $\phi(x) = c_0\phi(2x) + c_1\phi(2x-1) + c_2\phi(2x-2) + c_3\phi(2x-3) + \dots, k = 0, 1, 2, 3, \dots$, which says that $\phi(x)$ can be broken into a finite sequence of shifted copies of the compressed func-

tion $\phi(2x)$. The allowed shifts are $\frac{1}{2}k$, i.e., $0, \frac{1}{2}, 1\frac{1}{2}, \ldots$, and arbitrary coefficients c_k are allowed. (*a*) Show that $\Lambda(x-1)$ satisfies the dilation equation and give the coefficients c_k. (*b*) If $\Lambda(x-1)$ is represented by samples spaced $\Delta x = \frac{1}{2}$, i.e., by $\{0\ 1\ 2\ 1\ 0\}$, is the dilation equation satisfied? (*c*) Is the dilation equation satisfied by the Haar wavelet defined by $\phi(x) = \Pi(x)\,\mathrm{sgn}(-x)$?

3. **Slide whistle.** A sound lasts for one second during which the amplitude rises and then falls, while the pitch rises from 200 to 400 Hz and then falls. Write an expression that is consistent with the sound pressure signal during this occurrence. ▷

4. **Screeching halt.** A train approaching a road crossing is braking to a halt with an acceleration $-g$ and blowing its whistle. The fundamental frequency of the whistle, at rest, is 550 Hz. (*a*) Write an expression for the acoustic air pressure signal $s(t)$ at the crossing. ▷

5. **Mean frequency.** A signal $s(t)$ has a Fourier transform $S(f)$ and a power spectrum $|S(f)|^2$, which does not go negative, and in that respect resembles a mass distribution along a line. Define the mean frequency of a signal $s(t)$ as the centroid of the mass distribution that corresponds to the power spectrum, i.e., by

$$\frac{\int_{-\infty}^{\infty} f|S(f)|^2\,df}{\int_{-\infty}^{\infty} |S(f)|^2\,df}.$$

(*a*) Under this definition, what is the mean frequency for a signal $s(t) = \exp[-\pi(10^{-5}t)^2]\cos 2\pi 60t$? (*b*) What is the root-mean-square deviation of the power spectrum of this highly monochromatic signal from its mean frequency? (*c*) The waveform $s(t)$ is to be subjected to time-frequency analysis. What would be a suitable choice of cell size and shape? ▷

6. **A simple waveform.** A signal $s(t) = J_0(2\pi t)$, which approaches a frequency of 1 Hz as $|t| \to \infty$, would appear to have a lower frequency in the vicinity of $t = 0$, perhaps reaching about 2/3 Hz. How could this departure from monochromaticity be displayed on the (f, t)-diagram? ▷

7. **Fourier transform of Haar wavelet.** Derive the transform

$$F(s) = (\pi s)^{-1}(1 - \cos \pi s)(\sin \pi s + i \cos \pi s)$$

for the Haar wavelet defined by $f(x) = 1$ where $(0 < x < \frac{1}{2})$, $f(x) = -1$ where $(\frac{1}{2} < x < 1)$, and $f(x) = 0$ where $x < 0$ or $x > 1$. ▷

8. **Wigner distribution theorems.** If $f(t)$ has Wigner distribution $W(t, f)$, show that

$$f(-t) \quad \text{has W.D.} \quad W(-t, -f) \qquad (\pi \text{ rotation})$$

$$af(t) \quad \text{has W.D.} \quad |a|^2 W(t, f) \qquad (\text{nonlinearity})$$

$$f(at) \quad \text{has W.D.} \quad \frac{1}{|a|}\,W(at, f/a) \qquad (\text{similarity}).$$

9. **Wigner distribution for Gaussian pulse.** (*a*) Show that $f(t) = \exp(-\pi t^2/W^2)$ has Wigner distribution $W(t, f) = \sqrt{2}W \exp[-2\pi(t^2/W^2 - W^2 f^2)]$. (*b*) For a given equivalent width W, what is the shape of the loci $W(t, f) = \text{const}$?

20

Tables of sinc x, sinc^2 x, and exp $(-\pi x^2)$

In this table, the well-normalized functions sinc x, sinc^2 x, and exp $(-\pi x^2)$ are tabulated to six decimal places from 0 to 3.99. The definition of sinc x is

$$\text{sinc } x = \frac{\sin \pi x}{\pi x}.$$

Table of sinc x, sinc² x, and exp (−πx²)

x	sinc x	sinc^2 x	exp $(-\pi x^2)$	x	sinc x	sinc^2 x	exp $(-\pi x^2)$
.00	1.000 000	1.000 000	1.000 000	.25	.900 316	.810 569	.821 725
.01	.999 836	.999 671	.999 686	.26	.892 454	.796 473	.808 664
.02	.999 342	.998 685	.998 744	.27	.884 325	.782 031	.795 311
.03	.998 520	.997 043	.997 177	.28	.875 936	.767 263	.781 687
.04	.997 370	.994 747	.994 986	.29	.867 290	.752 192	.767 814
.05	.995 893	.991 802	.992 177	.30	.858 394	.736 840	.753 713
.06	.994 089	.988 212	.988 754	.31	.849 251	.721 228	.739 407
.07	.991 959	.983 983	.984 724	.32	.839 869	.705 379	.724 916
.08	.989 506	.979 121	.980 095	.33	.830 251	.689 316	.710 263
.09	.986 729	.973 634	.974 874	.34	.820 403	.673 061	.695 470
.10	.983 632	.967 531	.969 072	.35	.810 332	.656 638	.680 556
.11	.980 215	.960 821	.962 700	.36	.800 043	.640 068	.665 544
.12	.976 481	.953 515	.955 769	.37	.789 542	.623 376	.650 454
.13	.972 432	.945 623	.948 292	.38	.778 834	.606 583	.635 308
.14	.968 070	.937 159	.940 282	.39	.767 927	.589 712	.620 124
.15	.963 398	.928 135	.931 755	.40	.756 827	.572 787	.604 923
.16	.958 418	.918 566	.922 724	.41	.745 539	.555 828	.589 723
.17	.953 135	.908 466	.913 208	.42	.734 070	.538 859	.574 545
.18	.947 550	.897 851	.903 221	.43	.722 428	.521 902	.559 406
.19	.941 667	.886 736	.892 783	.44	.710 618	.504 977	.544 323
.20	.935 489	.875 140	.881 911	.45	.698 647	.488 107	.529 315
.21	.929 021	.863 080	.870 625	.46	.686 522	.471 312	.514 396
.22	.922 265	.850 573	.858 943	.47	.674 249	.454 612	.499 585
.23	.915 227	.837 640	.846 885	.48	.661 837	.438 028	.484 895
.24	.907 909	.824 298	.834 472	.49	.649 291	.421 579	.470 341

Table of sinc x, sinc² x, and exp $(-\pi x^2)$ (cont'd)

x	sinc x	sinc^2 x	exp $(-\pi x^2)$	x	sinc x	sinc^2 x	exp $(-\pi x^2)$
.50	.636 620	.405 285	.455 938	1.00	.000 000	.000 000	.043 214
.51	.623 829	.389 163	.441 698	1.01	−.009 899	.000 098	.040 570
.52	.610 926	.373 231	.427 634	1.02	−.019 595	.000 384	.038 063
.53	.597 919	.357 507	.413 758	1.03	−.029 083	.000 846	.035 689
.54	.584 815	.342 008	.400 081	1.04	−.038 360	.001 472	.033 442
.55	.571 620	.326 749	.386 613	1.05	−.047 423	.002 249	.031 317
.56	.558 342	.311 746	.373 363	1.06	−.056 269	.003 166	.029 308
.57	.544 989	.297 013	.360 341	1.07	−.064 895	.004 211	.027 411
.58	.531 568	.282 565	.347 555	1.08	−.073 297	.005 372	.025 621
.59	.518 086	.268 413	.335 012	1.09	−.081 473	.006 638	.023 932
.60	.504 551	.254 572	.322 719	1.10	−.089 421	.007 996	.022 341
.61	.490 970	.241 051	.310 682	1.11	−.097 138	.009 436	.020 843
.62	.477 350	.227 863	.298 905	1.12	−.104 623	.010 946	.019 432
.63	.463 699	.215 017	.287 395	1.13	−.111 873	.012 515	.018 106
.64	.450 024	.202 522	.276 154	1.14	−.118 886	.014 134	.016 860
.65	.436 333	.190 386	.265 186	1.15	−.125 661	.015 791	.015 690
.66	.422 632	.178 618	.254 494	1.16	−.132 196	.017 476	.014 591
.67	.408 929	.167 223	.244 080	1.17	−.138 490	.019 179	.013 561
.68	.395 232	.156 209	.233 944	1.18	−.144 541	.020 892	.012 596
.69	.381 548	.145 579	.224 089	1.19	−.150 350	.022 605	.011 692
.70	.367 883	.135 338	.214 514	1.20	−.155 915	.024 309	.010 847
.71	.354 245	.125 490	.205 219	1.21	−.161 235	.025 997	.010 056
.72	.340 642	.116 037	.196 204	1.22	−.166 310	.027 659	.009 317
.73	.327 079	.106 981	.187 467	1.23	−.171 140	.029 289	.008 627
.74	.313 565	.098 323	.179 006	1.24	−.175 724	.030 879	.007 982
.75	.300 105	.090 063	.170 820	1.25	−.180 063	.032 423	.007 382
.76	.286 708	.082 201	.162 906	1.26	−.184 157	.033 914	.006 822
.77	.273 379	.074 736	.155 261	1.27	−.188 006	.035 346	.006 301
.78	.260 126	.067 666	.147 881	1.28	−.191 611	.036 715	.005 816
.79	.246 955	.060 987	.140 764	1.29	−.194 972	.038 014	.005 365
.80	.233 872	.054 696	.133 906	1.30	−.198 091	.039 240	.004 945
.81	.220 885	.048 790	.127 301	1.31	−.200 968	.040 388	.004 556
.82	.207 999	.043 263	.120 947	1.32	−.203 604	.041 455	.004 195
.83	.195 220	.038 111	.114 837	1.33	−.206 002	.042 437	.003 860
.84	.182 556	.033 327	.108 967	1.34	−.208 162	.043 331	.003 549
.85	.170 011	.028 904	.103 333	1.35	−.210 086	.044 136	.003 262
.86	.157 593	.024 835	.097 928	1.36	−.211 776	.044 849	.002 995
.87	.145 306	.021 114	.092 748	1.37	−.213 234	.045 469	.002 749
.88	.133 156	.017 731	.087 786	1.38	−.214 462	.045 994	.002 522
.89	.121 150	.014 677	.083 038	1.39	−.215 462	.046 424	.002 311
.90	.109 292	.011 945	.078 497	1.40	−.216 236	.046 758	.002 117
.91	.097 589	.009 524	.074 158	1.41	−.216 788	.046 997	.001 939
.92	.086 044	.007 404	.070 015	1.42	−.217 119	.047 141	.001 774
.93	.074 664	.005 575	.066 062	1.43	−.217 234	.047 190	.001 622
.94	.063 452	.004 026	.062 293	1.44	−.217 133	.047 147	.001 482
.95	.052 415	.002 747	.058 702	1.45	−.216 821	.047 011	.001 353
.96	.041 557	.001 727	.055 283	1.46	−.216 301	.046 786	.001 235
.97	.030 882	.000 954	.052 031	1.47	−.215 576	.046 473	.001 126
.98	.020 395	.000 416	.048 939	1.48	−.214 650	.046 075	.001 027
.99	.010 099	.000 102	.046 002	1.49	−.213 525	.045 593	$.0_{3}9$ 353

Table of sinc x, sinc² x, and exp (−πx²) (cont'd)

x	sinc x	sinc² x	exp $(-\pi x^2)$	x	sinc x	sinc² x	exp $(-\pi x^2)$
1.50	−.212 207	.045 032	$.0_{3}8$ 514	**2.00**	.000 000	.000 000	$.0_{5}3$ 487
1.51	−.210 697	.044 393	$.0_{3}7$ 746	**2.01**	.004 974	.000 025	$.0_{5}3$ 075
1.52	−.209 001	.043 681	$.0_{3}7$ 043	**2.02**	.009 894	.000 098	$.0_{5}2$ 709
1.53	−.207 122	.042 900	$.0_{3}6$ 399	**2.03**	.014 756	.000 218	$.0_{5}2$ 385
1.54	−.205 065	.042 052	$.0_{3}5$ 811	**2.04**	.019 556	.000 382	$.0_{5}2$ 099
1.55	−.202 833	.041 141	$.0_{3}5$ 273	**2.05**	.024 290	.000 590	$.0_{5}1$ 846
1.56	−.200 431	.040 172	$.0_{3}4$ 782	**2.06**	.028 954	.000 838	$.0_{5}1$ 622
1.57	−.197 862	.039 150	$.0_{3}4$ 335	**2.07**	.033 545	.001 125	$.0_{5}1$ 425
1.58	−.195 133	.038 077	$.0_{3}3$ 926	**2.08**	.038 058	.001 448	$.0_{5}1$ 251
1.59	−.192 246	.036 958	$.0_{3}3$ 554	**2.09**	.042 491	.001 805	$.0_{5}1$ 097
1.60	−.189 207	.035 799	$.0_{3}3$ 215	**2.10**	.046 840	.002 194	$.0_{6}9$ 618
1.61	−.186 020	.034 603	$.0_{3}2$ 907	**2.11**	.051 101	.002 611	$.0_{6}8$ 427
1.62	−.182 690	.033 375	$.0_{3}2$ 626	**2.12**	.055 272	.003 055	$.0_{6}7$ 378
1.63	−.179 221	.032 120	$.0_{3}2$ 371	**2.13**	.059 350	.003 522	$.0_{6}6$ 456
1.64	−.175 619	.030 842	$.0_{3}2$ 140	**2.14**	.063 332	.004 011	$.0_{6}5$ 645
1.65	−.171 889	.029 546	$.0_{3}1$ 930	**2.15**	.067 214	.004 518	$.0_{6}4$ 934
1.66	−.168 034	.028 236	$.0_{3}1$ 739	**2.16**	.070 994	.005 040	$.0_{6}4$ 309
1.67	−.164 061	.026 916	$.0_{3}1$ 566	**2.17**	.074 670	.005 576	$.0_{6}3$ 761
1.68	−.159 975	.025 592	$.0_{3}1$ 410	**2.18**	.078 238	.006 121	$.0_{6}3$ 280
1.69	−.155 780	.024 267	$.0_{3}1$ 268	**2.19**	.081 697	.006 674	$.0_{6}2$ 860
1.70	−.151 481	.022 947	$.0_{3}1$ 140	**2.20**	.085 044	.007 233	$.0_{6}2$ 491
1.71	−.147 084	.021 634	$.0_{3}1$ 024	**2.21**	.088 278	.007 793	$.0_{6}2$ 169
1.72	−.142 594	.020 333	$.0_{4}9$ 197	**2.22**	.091 396	.008 353	$.0_{6}1$ 887
1.73	−.138 016	.019 048	$.0_{4}8$ 252	**2.23**	.094 396	.008 911	$.0_{6}1$ 641
1.74	−.133 355	.017 784	$.0_{4}7$ 400	**2.24**	.097 276	.009 463	$.0_{6}1$ 426
1.75	−.128 617	.016 542	$.0_{4}6$ 631	**2.25**	.100 035	.010 007	$.0_{6}1$ 238
1.76	−.123 806	.015 328	$.0_{4}5$ 939	**2.26**	.102 672	.010 541	$.0_{6}1$ 075
1.77	−.118 928	.014 144	$.0_{4}5$ 316	**2.27**	.105 184	.011 064	$.0_{7}9$ 322
1.78	−.113 988	.012 993	$.0_{4}4$ 755	**2.28**	.107 571	.011 572	$.0_{7}8$ 080
1.79	−.108 991	.011 879	$.0_{4}4$ 250	**2.29**	.109 832	.012 063	$.0_{7}7$ 000
1.80	−.103 943	.010 804	$.0_{4}3$ 797	**2.30**	.111 964	.012 536	$.0_{7}6$ 060
1.81	−.098 849	.009 771	$.0_{4}3$ 390	**2.31**	.113 969	.012 989	$.0_{7}5$ 243
1.82	−.093 714	.008 782	$.0_{4}3$ 024	**2.32**	.115 844	.013 420	$.0_{7}4$ 533
1.83	−.088 543	.007 840	$.0_{4}2$ 697	**2.33**	.117 589	.013 827	$.0_{7}3$ 917
1.84	−.083 341	.006 946	$.0_{4}2$ 403	**2.34**	.119 204	.014 210	$.0_{7}3$ 382
1.85	−.078 113	.006 102	$.0_{4}2$ 140	**2.35**	.120 688	.014 566	$.0_{7}2$ 919
1.86	−.072 865	.005 309	$.0_{4}1$ 905	**2.36**	.122 040	.014 894	$.0_{7}2$ 518
1.87	−.067 602	.004 570	$.0_{4}1$ 694	**2.37**	.123 262	.015 193	$.0_{7}2$ 170
1.88	−.062 329	.003 885	$.0_{4}1$ 506	**2.38**	.124 352	.015 463	$.0_{7}1$ 869
1.89	−.057 050	.003 255	$.0_{4}1$ 338	**2.39**	.125 310	.015 703	$.0_{7}1$ 609
1.90	−.051 770	.002 680	$.0_{4}1$ 187	**2.40**	.126 138	.015 911	$.0_{7}1$ 384
1.91	−.046 495	.002 162	$.0_{4}1$ 053	**2.41**	.126 834	.016 087	$.0_{7}1$ 190
1.92	−.041 229	.001 700	$.0_{5}9$ 340	**2.42**	.127 401	.016 231	$.0_{7}1$ 023
1.93	−.035 978	.001 294	$.0_{5}8$ 276	**2.43**	.127 837	.016 342	$.0_{8}8$ 780
1.94	−.030 745	.000 945	$.0_{5}7$ 329	**2.44**	.128 144	.016 421	$.0_{8}7$ 534
1.95	−.025 536	.000 652	$.0_{5}6$ 486	**2.45**	.128 323	.016 467	$.0_{8}6$ 461
1.96	−.020 354	.000 414	$.0_{5}5$ 736	**2.46**	.128 374	.016 480	$.0_{8}5$ 538
1.97	−.015 206	.000 231	$.0_{5}5$ 070	**2.47**	.128 298	.016 460	$.0_{8}4$ 743
1.98	−.010 094	.000 102	$.0_{5}4$ 478	**2.48**	.128 097	.016 409	$.0_{8}4$ 060
1.99	−.005 024	.000 025	$.0_{5}3$ 953	**2.49**	.127 772	.016 326	$.0_{8}3$ 473

Table of sinc x, sinc² x, and exp (−πx²) (cont'd)

x	sinc x	sinc² x	exp $(-\pi x^2)$	x	sinc x	sinc² x	exp $(-\pi x^2)$
2.50	.127 324	.016 211	$.0_{8}2\ 969$	**3.00**	.000 000	.000 000	$.0_{12}5\ 255$
2.51	.126 754	.016 067	$.0_{8}2\ 537$	**3.01**	−.003 322	.000 011	$.0_{12}4\ 351$
2.52	.126 064	.015 892	$.0_{8}2\ 166$	**3.02**	−.006 618	.000 044	$.0_{12}3\ 600$
2.53	.125 256	.015 689	$.0_{8}1\ 848$	**3.03**	−.009 886	.000 098	$.0_{12}2\ 977$
2.54	.124 331	.015 458	$.0_{8}1\ 576$	**3.04**	−.013 123	.000 172	$.0_{12}2\ 460$
2.55	.123 291	.015 201	$.0_{8}1\ 343$	**3.05**	−.016 326	.000 267	$.0_{12}2\ 032$
2.56	.122 137	.014 918	$.0_{8}1\ 144$	**3.06**	−.019 492	.000 380	$.0_{12}1\ 677$
2.57	.120 873	.014 610	$.0_{9}9\ 737$	**3.07**	−.022 618	.000 512	$.0_{12}1\ 383$
2.58	.119 500	.014 280	$.0_{9}8\ 283$	**3.08**	−.025 701	.000 661	$.0_{12}1\ 140$
2.59	.118 020	.013 929	$.0_{9}7\ 041$	**3.09**	−.028 740	.000 826	$.0_{13}9\ 393$
2.60	.116 435	.013 557	$.0_{9}5\ 982$	**3.10**	−.031 730	.001 007	$.0_{13}7\ 733$
2.61	.114 748	.013 167	$.0_{9}5\ 078$	**3.11**	−.034 670	.001 202	$.0_{13}6\ 362$
2.62	.112 961	.012 760	$.0_{9}4\ 309$	**3.12**	−.037 557	.001 411	$.0_{13}5\ 231$
2.63	.111 076	.012 338	$.0_{9}3\ 654$	**3.13**	−.040 389	.001 631	$.0_{13}4\ 299$
2.64	.109 097	.011 902	$.0_{9}3\ 096$	**3.14**	−.043 162	.001 863	$.0_{13}3\ 530$
2.65	.107 025	.011 454	$.0_{9}2\ 622$	**3.15**	−.045 876	.002 105	$.0_{13}2\ 897$
2.66	.104 864	.010 996	$.0_{9}2\ 219$	**3.16**	−.048 528	.002 355	$.0_{13}2\ 376$
2.67	.102 615	.010 530	$.0_{9}1\ 877$	**3.17**	−.051 114	.002 613	$.0_{13}1\ 948$
2.68	.100 283	.010 057	$.0_{9}1\ 587$	**3.18**	−.053 635	.002 877	$.0_{13}1\ 595$
2.69	.097 869	.009 578	$.0_{9}1\ 340$	**3.19**	−.056 087	.003 146	$.0_{13}1\ 306$
2.70	.095 377	.009 097	$.0_{9}1\ 132$	**3.20**	−.058 468	.003 419	$.0_{13}1\ 069$
2.71	.092 810	.008 614	$.0_{10}9\ 547$	**3.21**	−.060 777	.003 694	$.0_{14}8\ 736$
2.72	.090 170	.008 131	$.0_{10}8\ 050$	**3.22**	−.063 012	.003 971	$.0_{14}7\ 138$
2.73	.087 461	.007 649	$.0_{10}6\ 783$	**3.23**	−.065 171	.004 247	$.0_{14}5\ 829$
2.74	.084 685	.007 172	$.0_{10}5\ 712$	**3.24**	−.067 253	.004 523	$.0_{14}4\ 757$
2.75	.081 847	.006 699	$.0_{10}4\ 807$	**3.25**	−.069 255	.004 796	$.0_{14}3\ 879$
2.76	.078 949	.006 233	$.0_{10}4\ 043$	**3.26**	−.071 177	.005 066	$.0_{14}3\ 162$
2.77	.075 994	.005 775	$.0_{10}3\ 398$	**3.27**	−.073 018	.005 332	$.0_{14}2\ 575$
2.78	.072 985	.005 327	$.0_{10}2\ 855$	**3.28**	−.074 775	.005 591	$.0_{14}2\ 096$
2.79	.069 926	.004 890	$.0_{10}2\ 396$	**3.29**	−.076 448	.005 844	$.0_{14}1\ 706$
2.80	.066 821	.004 465	$.0_{10}2\ 010$	**3.30**	−.078 036	.006 090	$.0_{14}1\ 387$
2.81	.063 671	.004 054	$.0_{10}1\ 686$	**3.31**	−.079 537	.006 326	$.0_{14}1\ 127$
2.82	.060 482	.003 658	$.0_{10}1\ 412$	**3.32**	−.080 951	.006 553	$.0_{15}9\ 147$
2.83	.057 255	.003 278	$.0_{10}1\ 183$	**3.33**	−.082 277	.006 770	$.0_{15}7\ 423$
2.84	.053 995	.002 916	$.0_{11}9\ 897$	**3.34**	−.083 514	.006 975	$.0_{15}6\ 020$
2.85	.050 705	.002 571	$.0_{11}8\ 277$	**3.35**	−.084 662	.007 168	$.0_{15}4\ 879$
2.86	.047 388	.002 246	$.0_{11}6\ 917$	**3.36**	−.085 719	.007 348	$.0_{15}3\ 951$
2.87	.044 047	.001 940	$.0_{11}5\ 778$	**3.37**	−.086 686	.007 514	$.0_{15}3\ 198$
2.88	.040 687	.001 655	$.0_{11}4\ 823$	**3.38**	−.087 561	.007 667	$.0_{15}2\ 587$
2.89	.037 309	.001 392	$.0_{11}4\ 023$	**3.39**	−.088 346	.007 805	$.0_{15}2\ 091$
2.90	.033 918	.001 150	$.0_{11}3\ 354$	**3.40**	−.089 038	.007 928	$.0_{15}1\ 690$
2.91	.030 517	.000 931	$.0_{11}2\ 795$	**3.41**	−.089 640	.008 035	$.0_{15}1\ 364$
2.92	.027 110	.000 735	$.0_{11}2\ 327$	**3.42**	−.090 149	.008 127	$.0_{15}1\ 101$
2.93	.023 699	.000 562	$.0_{11}1\ 936$	**3.43**	−.090 567	.008 202	$.0_{16}8\ 877$
2.94	.020 288	.000 412	$.0_{11}1\ 610$	**3.44**	−.090 893	.008 262	$.0_{16}7\ 153$
2.95	.016 880	.000 285	$.0_{11}1\ 338$	**3.45**	−.091 128	.008 304	$.0_{16}5\ 761$
2.96	.013 478	.000 182	$.0_{11}1\ 111$	**3.46**	−.091 272	.008 331	$.0_{16}4\ 637$
2.97	.010 086	.000 102	$.0_{12}9\ 225$	**3.47**	−.091 325	.008 340	$.0_{16}3\ 730$
2.98	.006 707	.000 045	$.0_{12}7\ 652$	**3.48**	−.091 288	.008 333	$.0_{16}2\ 998$
2.99	.003 344	.000 011	$.0_{12}6\ 344$	**3.49**	−.091 161	.008 310	$.0_{16}2\ 409$

Table of sinc x, sinc² x, and exp (−πx²) (cont'd)

x	sinc x	sinc² x	exp $(-\pi x^2)$	x	sinc x	sinc² x	exp $(-\pi x^2)$
3.50	−.090 946	.008 271	$.0_{16}1$ 934	**3.75**	−.060 021	.003 603	$.0_{19}6$ 508
3.51	−.090 642	.008 216	$.0_{16}1$ 551	**3.76**	−.057 952	.003 358	$.0_{19}5$ 140
3.52	−.090 251	.008 145	$.0_{16}1$ 244	**3.77**	−.055 836	.003 118	$.0_{19}4$ 057
3.53	−.089 773	.008 059	$.0_{17}9$ 969	**3.78**	−.053 677	.002 881	$.0_{19}3$ 201
3.54	−.089 209	.007 958	$.0_{17}7$ 983	**3.79**	−.051 476	.002 650	$.0_{19}2$ 523
3.55	−.088 561	.007 843	$.0_{17}6$ 389	**3.80**	−.049 236	.002 424	$.0_{19}1$ 988
3.56	−.087 829	.007 714	$.0_{17}5$ 110	**3.81**	−.046 960	.002 205	$.0_{19}1$ 565
3.57	−.087 015	.007 572	$.0_{17}4$ 085	**3.82**	−.044 649	.001 994	$.0_{19}1$ 232
3.58	−.086 120	.007 417	$.0_{17}3$ 263	**3.83**	−.042 306	.001 790	$.0_{20}9$ 685
3.59	−.085 145	.007 250	$.0_{17}2$ 605	**3.84**	−.039 934	.001 595	$.0_{20}7$ 611
3.60	−.084 092	.007 071	$.0_{17}2$ 078	**3.85**	−.037 535	.001 409	$.0_{20}5$ 978
3.61	−.082 962	.006 883	$.0_{17}1$ 657	**3.86**	−.035 111	.001 233	$.0_{20}4$ 692
3.62	−.081 756	.006 684	$.0_{17}1$ 320	**3.87**	−.032 666	.001 067	$.0_{20}3$ 680
3.63	−.080 477	.006 476	$.0_{17}1$ 051	**3.88**	−.030 200	.000 912	$.0_{20}2$ 885
3.64	−.079 125	.006 261	$.0_{18}8$ 367	**3.89**	−.027 718	.000 768	$.0_{20}2$ 260
3.65	−.077 703	.006 038	$.0_{18}6$ 654	**3.90**	−.025 221	.000 636	$.0_{20}1$ 769
3.66	−.076 212	.005 808	$.0_{18}5$ 289	**3.91**	−.022 712	.000 516	$.0_{20}1$ 384
3.67	−.074 655	.005 573	$.0_{18}4$ 201	**3.92**	−.020 194	.000 408	$.0_{20}1$ 083
3.68	−.073 032	.005 334	$.0_{18}3$ 335	**3.93**	−.017 668	.000 312	$.0_{21}8$ 459
3.69	−.071 346	.005 090	$.0_{18}2$ 646	**3.94**	−.015 138	.000 229	$.0_{21}6$ 606
3.70	−.069 599	.004 844	$.0_{18}2$ 097	**3.95**	−.012 606	.000 159	$.0_{21}5$ 156
3.71	−.067 794	.004 596	$.0_{18}1$ 662	**3.96**	−.010 074	.000 101	$.0_{21}4$ 022
3.72	−.065 931	.004 347	$.0_{18}1$ 316	**3.97**	−.007 545	.000 057	$.0_{21}3$ 135
3.73	−.064 013	.004 098	$.0_{18}1$ 041	**3.98**	−.005 022	.000 025	$.0_{21}2$ 442
3.74	−.062 042	.003 849	$.0_{19}8$ 235	**3.99**	−.002 506	.000 006	$.0_{21}1$ 901

Half-peak abscissas

sinc 0.603355 = 0.5
$\text{sinc}^2\, 0.442946 = 0.5$
$\exp[-\pi(0.469719)^2] = 0.5$
$\text{sinc}\, 0.442946 = 1/\sqrt{2}$
$\exp[-\pi(0.332141)^2] = 1/\sqrt{2}$
$\text{sinc}'\, x = (\cos \pi x - \text{sinc}\, x)/x$

21

Solutions to Selected Problems

CHAPTER 2: GROUNDWORK

14. Since the real part of a complex quantity $f(x) = R(x) + iI(x)$ can be obtained by adding the quantity to its complex conjugate and dividing by 2,

$$R(x) = \frac{1}{2} f(x) + \frac{1}{2} f^*(x).$$

It is given that $f^*(x) = -f(-x)$. Hence

$$R(x) = \frac{1}{2} f(x) - \frac{1}{2} f(-x).$$

We see that $R(x)$ is odd because $R(x) + R(-x)$ is zero. For the imaginary part $I(x)$ we find

$$\begin{aligned} iI(x) &= \frac{1}{2} f(x) - \frac{1}{2} f^*(x) \\ &= \frac{1}{2} f(x) + \frac{1}{2} f(-x) \end{aligned}$$

which we see is even because $I(x) - I(-x) = 0$.

The Fourier transform of $f(x) = R(x) + iI(x)$ has two parts. The transform of $R(x)$ (which is real and odd) is imaginary by the previous problem. The transform of $I(x)$ (which is real and even) is real; therefore the transform of $iI(x)$ is imaginary and the whole of the transform is therefore imaginary.

19. The expression of 2.18 will be satisfactory as a measure of degree of symmetry, provided the integrals are evaluated over one period. For the first function mentioned, the even axes are at $x = 0, \pm 1, \pm 2, \ldots,$ and the degree of symmetry is 100 percent. The degree of odd symmetry would be 100 percent for $\Lambda(x) - \frac{1}{2}$ because that function becomes odd if shifted half a unit of x. The function $\Lambda(x) - \frac{1}{4}$ retains 100 percent even

symmetry but the odd symmetry is essentially destroyed, amounting only to 12.5 percent by the index.

$$\begin{aligned}&\int_{-1}^{1}[\Lambda(x)-\tfrac{1}{4}][\Lambda(x-1)-\tfrac{1}{4}]dx\Big/\int_{-1}^{1}[\Lambda(x)]^2\,dx\\&=2\int_0^1(\tfrac{3}{4}-x)(-\tfrac{1}{4}+x)\,dx\Big/\int_0^1(1-x)^2\,dx\\&=1/8.\end{aligned}$$

22. If $f(x)$ is even, then $f(x)+F(x)$ transforms into itself. For example, where $f(x)=1$ and $F(s)=\delta(s)$ then $1+\delta(x)$ transforms into itself; another example is $\Pi(x)+\operatorname{sinc}x$, which transforms into $\operatorname{sinc}x+\Pi(s)$. More generally (*J. Phys. A: Math. Gen.*, vol. 24, pp. L1143–1144, 1991), $f(x)+F(x)+f(-x)+F(-x)$ is self-transforming. This follows from the above because $f(x)+f(-x)$ is an even function whose transform is $F(s)+F(-s)$. The eigenfunctions of the Fourier transformation which are by definition self-transforming, are listed in Problem 8.25.

CHAPTER 3: CONVOLUTION

6. Let $f=g*h$. Then

$$\begin{aligned}f*f&=(g*h)*(g*h)\\&=g*[h*(g*h)]\\&=g*[h*(h*g)]\\&=g*[(h*h)*g]\\&=g*[g*(h*h)]\\&=(g*g)*(h*h),\text{ Q.E.D.}\end{aligned}$$

9. Theorem: $\int_{-\infty}^{\infty}f(g*h)\,dx=\int_{-\infty}^{\infty}g(h\star f)\,dx$

Derivation: $$\begin{aligned}\int_{-\infty}^{\infty}f(g*h)\,dx&=\int_{-\infty}^{\infty}f(x)\left[\int_{-\infty}^{\infty}g(u)\,h(x-u)\,du\right]dx\\&=\int_{-\infty}^{\infty}g(u)\left[\int_{-\infty}^{\infty}f(x)\,h(x-u)\,dx\right]du\\&=\int_{-\infty}^{\infty}g(u)[f*h(-)]\,du\\&=\int_{-\infty}^{\infty}g(h\star f)\,du.\end{aligned}$$

A more symmetrical theorem, not involving $f(g*h)$, but deducible from the foregoing by interchanging g and h, is

$$\int_{-\infty}^{\infty}g(h\star f)\,dx=\int_{-\infty}^{\infty}h(g\star f)\,dx.$$

Comment: A relation involving the Fourier transforms is

$$\int_{-\infty}^{\infty} f(g * h)\, dx = \int_{-\infty}^{\infty} F(-)\, GH\, ds.$$

11. Let $K_1(u)$ and $K_2(u)$ be two hermitian functions and let $S(u) = K_1(u) + K_2(u)$ and $P(u) = K_1(u)K_2(u)$. The sum $S(u)$ is hermitian by definition if $S(-u) = S^*(u)$. We see that

$$\begin{aligned} S(-u) &= K_1(-u) + K_2(-u) = K_1^*(u) + K_2^*(u) = [K_1(u) + K_2(u)]^* \\ &= S^*(u), \end{aligned}$$

so the sum is hermitian. Now

$$\begin{aligned} P(-u) &= K_1(-u)K_2(-u) = K_1^*(u)K_2^*(u) = [K_1(u)K_2(u)]^* \\ &= P^*(u) \end{aligned}$$

and the product is also hermitian.

17. Let $x = 0.2\,n$ where n is integral. Then the problem is first to find approximate values for $f_1(x)$ where

$$\{f_1(n)\} * \{1.0 \;\; 0.8187 \;\; 0.6703 \;\; 0.5488 \;\; 0.4493 \;\; 0.3679 \;\ldots\} = \{1\}.$$

By the method of Fig. 3.10 we find $\{f_1(n)\} = \{1 \;\; -0.8187\}$. The normalized sequence asked for is $\{1 \;\; -0.8187\}/0.1812 = \{5.518 \;\; -4.518\}$.

25. The desired function is given by the area of overlap of two unit-diameter circles whose centers are separated by r. If the centers are at C and C', Q is the midpoint of CC', P is a point of intersection of the two circles and R is the point nearest to C' on the circle with center at C, the area is four times the segment PQR.

$$\begin{aligned} \text{chat}\, r &= 4(\text{sector } CPR - \text{triangle } CPQ) = 4\left[\frac{1}{2} CP^2 \cos^{-1} r - \frac{1}{2} CQ \cdot PQ\right] \\ &= \frac{1}{2}[\cos^{-1} r - r(1 - r^2)^{1/2}]. \end{aligned}$$

29. (*a*) Each sequence is the reciprocal of the other as verified by taking the convolution sum. (*b*) Two functions whose samples agree with the two sequences are $v(t) = 2H(t)$ and $i(t) = 2 \cos \pi t\, H(t)$.

30. Any sequence having period 2, 3, or 6 elements is invisible to convolution with the 6-element $\{a_k\}$; consequently no periodic components of period 2, 3, or 6 will be present in the given $\{c_k\}$. The deconvolution algorithm cannot recreate them; it will produce that part of the unknown $\{b_k\}$ from which the components of period 2, 3, and 6 have been removed—known as the principal solution (*Austral. J. Phys.*, vol. 7, pp. 615–640, 1954). In some circumstances the unknown $\{b_k\}$ may nevertheless be recoverable, for example where $\{b_k\}$ has finite extent. This important subject comes under the heading of restoration.

31. Time to convolve should not depend on which sequence is shorter. Asymptomatic variation with N will depend on whether the convolution is computed by direct multiplication and adding, by the matrix multiplication method of Chapter 3, by multiplication of transforms, or by a built-in routine. The interesting thing about timing is that the asymptotic variation for your chosen method can be determined empirically.

CHAPTER 4: NOTATION FOR SOME USEFUL FUNCTIONS

7. $$4 \operatorname{sinc} 4x * \sin x = \int_{-\infty}^{\infty} \frac{\sin(x-u)\sin 4\pi x}{\pi u}\,du$$
$$= \frac{\sin x}{\pi}\int_{-\infty}^{\infty}\frac{\cos u \sin 4\pi u}{u}\,du - \frac{\cos x}{4\pi}\int_{-\infty}^{\infty}\frac{\sin u \sin 4\pi u}{u}\,du$$
$$= \frac{2\sin x}{\pi}\int_{0}^{\infty}\frac{\cos u \sin 4\pi u}{u}\,du - \text{(null term due to odd integrand).}$$

We need the standard integral (e.g., see Gradshteyn and Ryzhik No. 3.741.2).

$$\int_0^{\infty}\frac{\sin ax \cos bx}{x}\,dx = \pi/2, \qquad a > b \ge 0.$$

Thus

$$\int_0^{\infty}\frac{\sin 4\pi u \cos u}{u}\,du = \pi/2$$

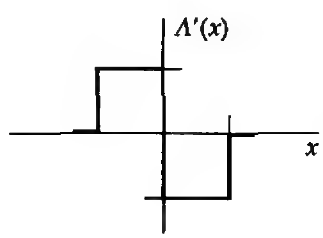
and $4 \operatorname{sinc} 4x * \sin x = \sin x$.

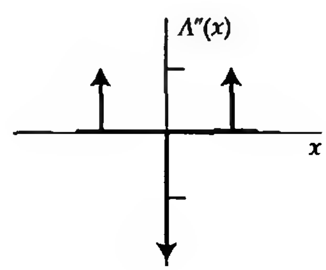
12. We find that

$$\Lambda'(x) = \begin{cases} 0 & |x| > 1 \\ 1 & -1 < x < 0 \\ -1 & 0 < x < 1 \end{cases}$$

and we see that this is equal to $-\Pi(x/2)\operatorname{sgn} x$. The second derivative, away from the points of discontinuity at $x = -1, 0, 1$, is zero. In classical analysis we say that $\Lambda'(x)$ is not differentiable at those points, or that the second derivative of $\Lambda(x)$ does not exist at those points. However, in the next chapter such a situation will be handled by delta-symbol notation, as follows.

$$\Lambda''(x) = \delta(x+1) - 2\delta(x) + \delta(x-1).$$





15. By definition

$$|x| = \begin{cases} x & x > 0 \\ -x & x < 0 \end{cases}$$

and, because $|x|$ is continuous,

$$\frac{d}{dx}|x| = \begin{cases} 1 & x > 0 \\ -1 & x < 0 \end{cases}$$
$$= \operatorname{sgn} x.$$

Now, sgn x is not continuous; therefore we cannot simply differentiate the expressions 1 and -1 in the two ranges of x and conclude that $(d/dx)\operatorname{sgn} x$ is zero everywhere. The discontinuity of amount $+2$ at $x = 0$ will produce a result $\operatorname{sgn}' x = 2\delta(x)$ in the notation of the next chapter. The fact that $|x|$ and $2xH(x)$ have the same second derivative is due to the fact that their first derivatives differ by a constant as may be seen by differentiating $|x| - x = 2xH(x)$.

19. The result can be confirmed with a lengthy table of integrals or, equivalently, from a table of cosine transforms, both of which are available as software packages. Also the Pictorial Dictionary shows that $\operatorname{jinc} x \supset (1 - s^2)\Pi(s/2)$. The desired infinite integral equals the value of the transform at $s = 0$, which is unity.

CHAPTER 5: THE IMPULSE SYMBOL

3. The expression $\delta(2x^2 - \frac{1}{2})$ is zero except where $2x^2 - \frac{1}{2} = 0$, i.e., at $x = \pm\frac{1}{2}$ where there will be impulses. The strength of each impulse will be the reciprocal of the absolute slope of $2x^2 - \frac{1}{2}$, which is 2 at both $x = \frac{1}{2}$ and $x = \frac{1}{2}$, Hence $\delta(2x^2 - \frac{1}{2}) = \frac{1}{2}\delta(x - \frac{1}{2}) + \frac{1}{2}\delta(x + \frac{1}{2})$. In the case of $\delta(x^2 - a^2)$ the absolute slope is $2a$ at $x = a$ and $x = -a$. Hence $\delta(x^2 - a^2) = \delta(x - a)/2a + \delta(x + a)/2a$.

7. (*a*) The roots of $\sin \pi x$ are at $x = 0, \pm 1, \pm 2, \ldots$ and the absolute slope at the zero crossings is π. Therefore $\delta(\sin \pi x)$ consists of impulses of strength π^{-1} at integral values of x. Consequently $\pi\delta(\sin \pi x)$ is the same as $\mathrm{III}(x)$.

   (*b*) In the case of $\delta(\sin x)$ the impulses are of *unit* strength situated at $x = 0, \pm\pi, \pm 2\pi, \ldots$ and the ensemble is represented by $\pi^{-1}\,\mathrm{III}(x/\pi)$. The π in parentheses stretches the abscissa to make the impulse separation π instead of unity as for $\mathrm{III}(x)$. But the stretching strengthens the impulses; hence the compensating factor π^{-1}.

9. A graph will show that

$$\mathrm{III}(x)\,\Pi(x/8) = \frac{1}{2}\delta(x + 4) + \mathrm{III}(x)\,\Pi(x/7)\frac{1}{2}\delta(x - 4)$$

provided we take $\Pi(\pm\frac{1}{2})$ to be $\frac{1}{2}$. This feature was not provided for in the definition of $\Pi(x)$. It arises here, not out of the definition of $\Pi(x)$, but out of the sifting integral when the sifting impulse falls at a discontinuity of the function being sampled. The result is (p. 76)

$$\int_{-\infty}^{\infty} \delta(x - a)\,f(x) = [f(a+) + f(a-)]/2.$$

In our example we could write, for the right-hand edge of $\Pi(x/8)$,

$$\int_{3.5}^{4.5} \delta(x-4)\,\Pi(x/8)\,dx = \frac{1}{2}.$$

Consequently $\delta(x-4)\,\Pi(x/8)$ is equivalent to $\frac{1}{2}\delta(x-4)$ and similarly for the left-hand edge where $\delta(x+4)\,\Pi(x/8) = \frac{1}{2}\delta(x+4)$.

12. (*a*) A satisfactory proof that $\delta'(-x) = -\delta'(x)$ is as follows. The sequence $p_\tau(x) = \tau^{-1}\exp(-\pi x^2/\tau^2)$ defines $\delta(x)$ and

$$\int_{-\infty}^{\infty} \delta(x)\,F(x)\,dx \triangleq \lim_{\tau\to 0} \int_{-\infty}^{\infty} p_\tau(x)\,F(x)\,dx = F(0) \tag{1}$$

in the notation of p. 92. The sequence $p_\tau(x)$ defines $\delta'(x)$. Let us now establish that

$$\int_{-\infty}^{\infty} \delta'(x)\,F(x)\,dx = -F'(0). \tag{2}$$

By definition

$$\begin{aligned}\int_{-\infty}^{\infty} \delta'(x)\,F(x)\,dx &= \lim_{\tau\to 0} \int_{-\infty}^{\infty} p'_\tau(x)\,F(x)\,dx \\ &= -\lim_{\tau\to 0} \int_{-\infty}^{\infty} p_\tau(x)\,F'(x)\,dx \qquad \text{(p. 93)} \\ &= -F'(0) \qquad \text{from (1).}\end{aligned}$$

Now for $\delta'(-x)$:

$$\begin{aligned}\int_{-\infty}^{\infty} \delta'(-x)\,F(x)\,dx &= \lim_{\tau\to 0} \int_{-\infty}^{\infty} p'_\tau(-x)\,F(x)\,dx \\ &= -\lim_{\tau\to 0} \int_{-\infty}^{\infty} p'_\tau(x)\,F(x)\,dx \qquad (p'_\tau \text{ being odd}) \\ &= F'(0).\end{aligned}$$

This discussion has provided an opportunity to exhibit the derivation of the basic relations (1) and (2) using generalized functions. But the argument can be summed up as follows. The generalized function $\delta'(-x)$ is the negative of $\delta'(x)$ because the sequence defining $\delta'(-x)$ is the negative of the one defining $\delta'(x)$.

An interesting feature of the result is that there are sequences equivalent to $p_\tau(x)$ that are not even and sequences equivalent to $p'_\tau(x)$ that are not odd. Nevertheless $\delta(x)$ and $\delta'(x)$ possess the properties of evenness and oddness respectively, in generalized function theory. For more on this see 5.19.

In another approach we may regard

$$\delta'(x) * f(x) = \frac{d}{dx} f(x) \tag{3}$$

as a basic property of $\delta'(x)$. It follows by reversing both sides that

$$\delta'(x) * f(-x) = \frac{d}{d(-x)} f(-x).$$

Let $g(x) = f(-x)$. Then

$$\delta'(-x) * g(x) = -\frac{d}{dx} g(x). \tag{4}$$

Comparison of (3) and (4) shows that $\delta'(-x)$ behaves like $-\delta'(x)$.

(*b*) To investigate $x\delta'(x)$, multiply it by a test function $F(x)$ and integrate. From (2)

$$\int_{-\infty}^{\infty} x\delta'(x)\,F(x)\,dx = -\frac{d}{dx}[xF(x)]_{x=0}$$
$$= -F(0).$$

But
$$\int_{-\infty}^{\infty}[-\delta(x)]F(x)\,dx = -F(0). \text{ Hence } x\delta' = -\delta(x).$$

(*c*)
$$\frac{d}{dx}[f(x)\,\delta(x)] = \frac{d}{dx}[f(0)\,\delta(x)] = f(0)\,\delta'(x).$$

Also

$$\frac{d}{dx}[f(x)\,\delta(x)] = f(x)\,\delta'(x) + f'(x)\,\delta(x)$$
$$= f(x)\,\delta'(x) + f'(0)\,\delta(x).$$

Hence
$$f(x)\,\delta'(x) = f(0)\,\delta'(x) - f'(0)\,\delta(x).$$

All the foregoing discussions can be redone in terms of the sequence $\tau^{-1}\Pi(x/\tau)$ and its derivative $\tau^{-1}[\delta(x+\tau/2) - \delta(x-\tau/2)]$ with a view to reinforcing intuitive feeling for the results.

19. The "causal" sequence $\tau^{-1}\Pi[(x-\frac{1}{2}\tau)/\tau]$ $(\tau > 0)$ is sometimes adopted in books dealing with the response of systems whose inputs switch on at $t = 0$. The impulse response is then the limit of the response to rectangular pulses as $\tau \to 0$. The sequence of rectangular functions defines an entity $\delta_+(x)$ with the property

$$G(0+) = \lim_{\tau\to 0}\int_{-\infty}^{\infty}\tau^{-1}\Pi[(x-\tfrac{1}{2}\tau)/\tau]G(x)\,dx = \int_{-\infty}^{\infty}\delta_+(x)\,G(x)\,dx$$

where
$$G(0+) = \lim_{\epsilon\to 0} G(x + |\epsilon|).$$

A similar property holds for $\delta_-(x)$. Since (p. 76)

$$\int_{-\infty}^{\infty}\delta(x)\,G(x)\,dx = \frac{1}{2}G(0-) + \frac{1}{2}G(0+),$$

it follows that $\delta(x) = \frac{1}{2}\delta_-(x) + \frac{1}{2}\delta_+(x)$.

The asymmetrical profile $\tau^{-1}\{\Lambda(x/\tau) + \frac{1}{2}\Lambda[(x-\tau)/\tau]\}$ contains two parts, the first behaving as $\delta(x)$, the second as $\frac{1}{2}\delta_+(x)$. It defines an entity $\delta_a(x) = \delta(x) + \frac{1}{2}\delta_+(x) = \delta_+(x) + \frac{1}{2}\delta_-(x)$ (i.e., $\mu = 1, \nu = \frac{1}{2}$).

In the theory of generalized functions many problems are circumvented by limiting attention to sequences and test functions that have derivatives of all orders. Since we are always applying impulse symbol notation to test functions $G(x)$ that do not meet the requirements, subsidiary arguments are often required to establish that a standard result will remain true even though $G(x)$ is defective. In addition, however, there are phenomena that arise in practice that are not in the theory of generalized functions at all. The notion of $\delta_+(x)$, which sifts out the value at the right-hand side of a discontinuity is one of these.

24. (*a*) When we take the limit of $\frac{1}{2}\tau^{-1}\Lambda[(x/\tau)-1]$ as $\tau \to 0$ we must first fix x. If we fix x at 0, the function value is zero for all τ and so the limit is zero. If $x \neq 0$, the function value rises to a peak and declines as $\tau \to 0$; for $\tau < x/2$ it remains zero and thus has a limit of zero. The function $f(x,\tau)$ quoted has unit area but approaches zero for all x, in the limit as $\tau \to 0$. (One might note that $\max[f(x,\tau)]$ increases without limit as $\tau \to 0$.) Regular sequences (p. 90) also exist with the same properties. Nevertheless such sequences are equivalent to other sequences for $\delta(x)$.

(*b*) An example where $f(0,\tau) \to -\infty$ as $\tau \to 0$ is furnished by $\tau^{-1}[\Lambda(\tau^{-1}x-2) - \Lambda(\tau^{-1}x) + \Lambda(\tau^{-1}x+2)]$, representing a negative central triangle function flanked by two equal positive ones, the whole having unit area.

25. (*a*) From Problem 5.12, $f(x)\delta'(x) = f(0)\delta'(x) - f'(0)\delta(x)$.
Differentiate to obtain $f(x)\delta''(x) + f'(x)\delta'(x) = f(0)\delta''(x) - f'(0)\delta'(x)$.
Thus $f(x)\delta''(x) = f(0)\delta''(x) - f'(0)\delta'(x) - f'(x)\delta'(x)$. Using 5.12 again to expand the third term we have

$$f(x)\delta''(x) = f(0)\delta''(x) - 2f'(0)\delta'(x) + f''(0)\delta(x).$$

(*b*) Prove this part by mathematical induction. If $f(x)\delta^{(n)}(x) =$

$$a_0 f(0)\delta^{(n)}(x) - a_1 f'(0)\delta^{(n-1)}(x) + \ldots + a_n f^{(n)}(0)\delta(x) \tag{1}$$

then

$$f(x)\delta^{(n+1)}(x) + f'(x)\delta^{(n)}(x) = a_0 f(0)\delta^{(n+1)}(x) - a_1 f'(0)\delta^{(n)}(x) + \ldots + a_n f^{(n)}(0)\delta'(x)$$

and, transposing and using (1),

$$\begin{aligned} & f(x)\delta^{(n+1)}(x) \\ &= a_0 f(0)\delta^{(n+1)}(x) - a_1 f'(0)\delta^{(n)}(x) + \ldots + a_n f^{(n)}(0)\delta'(x) \\ &\qquad - a_0 f'(0)\delta^{(n)}(x) + \ldots + a_{n-1} f^{(n)}(0)\delta'(x) + a_n f^{(n+1)}(0)\delta(x) \\ &= b_0 f(0)\delta^{(n+1)}(x) - b_1 f'(0)\delta^{(n)}(x) + \ldots + b_n f^{(n)}(0)\delta'(x) - b_{n+1} f^{(n+1)}(0)\delta(x). \end{aligned} \tag{2}$$

We see that the sequence $\{b_0\, b_1 \cdots b_n\} = \{1\ 1\} * \{a_0\, a_1 \cdots a_n\}$. Now (1) is true for $n = 2$, with binomial coefficients $\{1\ 2\ 1\}$. Therefore (2) is true for $n = 3$, with binomial coefficients $\{1\ 3\ 3\ 1\}$ and so for $n = 4$ and for all integral n.

32. As the waveform should not matter let us select a rectangular pulse. We can also study the current $I(t)$. Thus $V(t) = K\Pi(t/\tau)$ and since $V(t) = R\,I(t)$ it follows that $I(t) = (K/R)\Pi(t/\tau)$. It is given that

$$\int V(t)I(t)dt = 2.$$

Now $V(t)I(t) = (K^2/R)\Pi(t/\tau)$. Therefore

$$\int (K^2/R)\Pi(t/\tau)dt = 2$$

from which we find $K^2 = 2R/\tau$. Thus for the instantaneous power

$$V(t)I(t) = 2\tau^{-1}\Pi(t/\tau)$$

which tells us that the power is describable by an impulse $2\delta(t)$ as τ becomes too brief to worry about. However, if the voltage itself were a regular impulse, infinite energy

would be transferred to the resistor, contrary to what is given. Noting that

$$V(t)I(t) = [V(t)]^2/R = 2R^{-1}\tau^{-1}\,\Pi(t/\tau)$$

we obtain

$$V(t) = 2^{1/2}\tau^{-1/2}\,\Pi(t/\tau).$$

The desired δ-notation for $V(t)$ is therefore $2^{1/2}\,\delta^{1/2}(t/\tau)$. Even though the voltage waveform is a null function, it can deliver finite energy to a resistor.

33. (*a*) $$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\delta(x)\delta(y)\,dx\,dy = \lim_{\tau\to 0}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\tau^{-1}\Pi(x/\tau)\tau^{-1}\Pi(y/\tau)\,dx\,dy$$
$$= \lim_{\tau\to 0}\int_{-\tau/2}^{\tau/2}\int_{-\tau/2}^{\tau/2}\tau^{-2}\,dx\,dy = 1.$$

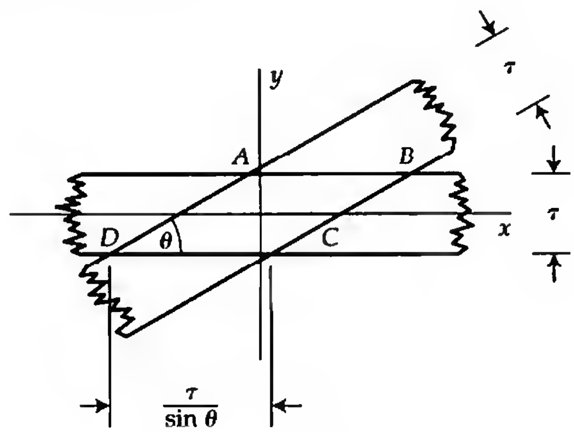
Because of this property we write $\delta(x)\delta(y) = {}^2\delta(x,y)$ as on p. 85.

(*b*) A unit blade making an angle θ with the x-axis is described by $\delta(y - x\tan\theta)$ and $\delta(y)$ is a unit blade along the x-axis. Thus the problem asks us to consider an integral such as

$$I = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\delta(y - x\tan\theta)\delta(y)\,dx\,dy.$$

Following the rule,

$$I = \lim_{\tau\to 0}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\tau^{-1}\Pi\left(\frac{y - x\tan\theta}{\tau}\right)\tau^{-1}\Pi(y/\tau)dx\,dy$$
$$= \lim_{\tau\to 0}\iint_{ABCD}\tau^{-2}\,dx\,dy.$$



The integrand is τ^{-2} over the parallelogram ABCD whose height is τ, base $\tau/\sin\theta$ and area $\tau^2/\sin\theta$. Hence the value of the integral is $1/\sin\theta$ and approaches infinity as the intersection becomes more and more skew.

35.
$$\int_{-\infty}^{\infty}\delta(\sin x)\Pi\left(x - \frac{1}{2}\right)dx = \frac{1}{2}$$
$$\int_{-\infty}^{\infty}\delta(\cos x)\,\Pi(x/4)\,dx = 2$$
$$\int_{-\infty}^{\infty}\delta(\sin 2x)\Pi(x/4)\,dx = \frac{3}{2}.$$

36. There is a strong objection from practitioners of the single-sided Laplace transform, which they can avoid by changing their notation to $\delta_+(x)$ for explanatory purposes (see Problem 5.19).

37. Because $|x|$ is continuous its derivative, away from the origin of x, simply equals $\operatorname{sgn}(x)$. So the asserted equality is subject to the condition $x \neq 0$. At $x = 0$ the derivative will be an undetermined constant but it disappears on further differentiation. As $|x| + x$ is identically equal to $2xH(x)$, both these functions have the same second derivative $2\delta(x)$, even though both are undefined at the origin.

38. Examine the normalized sum $N^{-1}\Sigma_1^N 2 \sin 2\pi ks$ over the range $0 < s < 0.25$ for larger and larger values of N, such as $N = 200, 400$, and 800. The graph will show, if it was not otherwise obvious, that $\Sigma_k 2 \sin 2\pi ks$ does not converge, but will also show that the partial sums oscillate evenly about the value $\frac{1}{2} \cos \frac{1}{2}^\circ$ when s is chosen so that $2\pi s = 1^\circ$. In problems where $\text{III}(x) \operatorname{sgn} x$ occurs as a factor in an expression that is physically more realistic, adoption of the suggested transform can lead to a rigorously correct result.

CHAPTER 6: THE BASIC THEOREMS

12. Let the notation $f^{*3/2}$ mean f convolved with itself half a time. From the previous problem it would be reasonable to require that

$$f^{*3/2} \supset F^{3/2}$$

or

$$f^{*3/2} \equiv \int_{-\infty}^{\infty} e^{i2\pi sx} \left[\int e^{-i2\pi su} f(u)\, du \right]^{3/2} ds.$$

This integral expression provides an operational definition for fractional-order self-convolution. For a different view see Chapter 13.

14. From Problem 6.1(d),

$$e^{-ax^2} \supset (\pi/a)^{\frac{1}{2}} e^{-\pi^2 s^{2/a}}$$

Hence

$$e^{-ax^2} * e^{-bx^2} \supset \pi\,(ab)^{-\frac{1}{2}} e^{-\pi^2 s^2 (a^{-1}+b^{-1})}.$$

20. Let $g(x) = \int^x f(u)\, du$. Then $\int_{-\infty}^{\infty} G(s)\, ds = g(0)$. Consequently

$$\int_{-\infty}^{\infty} \mathcal{F}\left[\int^x f(u)\, du \right] dx = \int^0 f(u)\, du.$$

26.
$$\int_{-\infty}^{\infty} \operatorname{sinc}^2 x \cos \pi x \, dx = \int_{-\infty}^{\infty} \Lambda(s)\left[\frac{1}{2}\delta(s + \tfrac{1}{2}) + \frac{1}{2}\delta(s - \tfrac{1}{2})\right]ds$$
$$= \frac{1}{2}.$$

28. First verify that $(\pi x)^{-1} \supset -i \operatorname{sgn} s$, which is easily done by evaluating the Fourier integral of the R.H.S. Then

$$(\pi x)^{-1} * (-\pi x)^{-1} \supset (-i \operatorname{sgn} s)(i \operatorname{sgn} s) = \operatorname{sgn}^2 s = 1.$$

The function whose transform is 1 is $\delta(x)$. It follows that any function whose transform jumps between 1 and -1 will have an impulsive autocorrelation. Examples of such functions are $2 \operatorname{sinc} x - \delta(x)$, $\delta(x) - 4 \operatorname{sinc} x \cos \omega x$ $(\omega > 2\pi)$, and $(\pi x)^{-1} e^{ix}$.

29. The surprising result seems to say that

$$e^{-\pi x^2} = \delta(x) + (4\pi)^{-1}\delta''(x) + \ldots, \tag{1}$$

which cannot possibly be true. The flaw appears if we write the reasoning out in full as follows:

$$f(x) = \mathcal{F} \lim_{N\to\infty} \sum_{n=0}^{N} \frac{s^n}{n!} F^{(n)}(0) \tag{2}$$

$$= \lim_{N\to\infty} \sum_{n=0}^{N} \mathcal{F}[s^n F^{(n)}(0)/n!] \tag{3}$$

$$= \lim_{N\to\infty}\left[\delta(x)\int_{-\infty}^{\infty} f(x)\,dx - \ldots + \delta^{(N)}(x)\int_{-\infty}^{\infty} \frac{x^N}{N!} f(x)\,dx\right].$$

We know that the Fourier transform of a sum of a finite number of functions is the sum of the separate transforms, but is the transform of a limit of a sum (2) equal to the limit of the sum of the transforms (3)? Evidently not. The situation is analogous to a doubly infinite series that has to be summed over rows and columns. If we plan to obtain the row sums first, then sum those, the row sums themselves must exist. In the present case we move from (2), where the sum for each s does exist in the limit, to a rearrangement (3), where Fourier integrals over s do not in the strict sense exist. Even so, the peculiar relation (1) may have some application. For example, if we had to find the transform of $e^{-\pi s^2}\Pi(s/a)$, which is not analytically convenient, would a correct series result be furnished by

$$a \operatorname{sinc} ax * [\delta(x) + (4\pi)^{-1}\delta''(x) + \ldots]?$$

33. Let (Exp) mean any function of the form $a\exp(-|x|)$. Then (Voigt) = (Gaussian) $*$ (Cauchy) $\supset$ (Gaussian) $\times$ (Exp).

So (Voigt) $*$ (Voigt) $\supset$ (Gaussian) $\times$ (Gaussian) $\times$ (Exp) $\times$ (Exp)
= (Gaussian) $\times$ (Exp).

Hence (Voigt) $*$ (Voigt) = (Voigt).

CHAPTER 7: OBTAINING TRANSFORMS

1. The second derivative $F''(s)$ is approximated by the second finite difference $F(s + \Delta s) - 2F(s) + F(s - \Delta s)$ divided by $(\Delta s)^2$. So $F''(0)$ is approximated by $[F(\Delta s) - 2F(0) + F(-\Delta s)]/(\Delta s)^2$. The imaginary part of $F(s)$, being odd, has no curvature at $s = 0$ while the real part, being even, allows $F(-\Delta s)$ to be replaced by $F(\Delta s)$. Consequently, the expression $-F''(0)/4\pi^2$ is approximated by $2[R(0) - R(1)]/4\pi^2(\Delta s)^2$, a quantity that should be a little less than the second moment of the data (the agreement improving as Δs is smaller). After $R(0)$ has been verified as being equal to $\sum f(x)$, the moment test checks that $R(1)$ has no gross error.

2. See tables of integrals under trigonometrical functions of complicated arguments, or tables of cosine transforms. A simple numerical check is to sum $\cos \pi x^2$ at intervals Δx, multiply by Δx, and compare with $F(0) = 2^{-1/2}$. The summand becomes so highly oscillatory as x increases that numerical summing might seem unreliable. However, most of the contribution to the integral comes in the first few cycles around $x = 0$, the rapid oscillations at larger $|x|$ tending to cancel. Thus the sum from $x = 0$ to 10 with $\Delta x = 0.1$ agrees with expectation within a few percent, immediately eliminating gross errors. Repeating for $x = 0$ to 100 with $\Delta x = 0.001$ gives agreement to 0.2 percent. While numerical approximation of this kind does not constitute a proof, it gives fast reassurance against the factors of 2, π, and so on, that tend to creep into algebra and calculus.

3. It is already known that $\cos \pi x^2 \supset 2^{-1/2}(\cos \pi s^2 + \sin \pi s^2)$. The additional integral $i\int_{-\infty}^{\infty}\sin \pi x^2 e^{-i2\pi sx}dx$ equals $2i\int_0^{\infty}\sin \pi x^2 \cos 2\pi sx\, dx$ (after omitting the infinite integral of $\sin \pi x^2 \sin 2\pi sx$ on the grounds that this integrand is an odd function of x). From tables of integrals or of cosine transforms we conclude that

$$\begin{aligned} e^{i\pi x^2} &\supset 2^{-1/2}[\cos \pi s^2 + \sin \pi s^2 + i(\cos \pi s^2 - \sin \pi s^2)] \\ &= 2^{-1/2}[e^{-i\pi s^2} + ie^{-i\pi s^2}] = \sqrt{i}e^{-i\pi s^2}. \end{aligned}$$

While tables are convenient, transforms can also be worked out by hand. Take this case as an example. Since neither $\cos x^2$ nor $\sin x^2$ dies out as x increases, introduce damping that will later be allowed to approach zero. Thus consider $\exp(-\sigma x^2)\exp ix^2$, where σ will later approach zero. Then

$$\begin{aligned} \Phi(s) &= \int_{-\infty}^{\infty} e^{-(\sigma - i)x^2}e^{-i2\pi sx}dx \\ &= e^{-\pi^2 s^2/(\sigma - i)}\int_{-\infty}^{\infty} e^{-(\sigma - i)x^2 - i2\pi sx + \pi^2 s^2/(\sigma - i)}dx \\ &= e^{-\pi^2 s^2/(\sigma - i)}\int_{-\infty}^{\infty} e^{-(\sigma - i)[x - i\pi s/(\sigma - i)]^2}dx \\ &= e^{-\pi^2 s^2/(\sigma - i)}\int_{-\infty}^{\infty} e^{-(\sigma - i)x^2}dx = e^{-\pi^2 s^2/(\sigma - i)}\sqrt{\frac{\pi}{\sigma - i}}. \end{aligned}$$

Finally

$$e^{ix^2} \supset \lim_{\sigma \to 0}\Phi(s) = \sqrt{\frac{\pi}{-i}}e^{-i\pi^2 s^2}.$$

From this it follows that

$$e^{i\pi x^2} \supset \sqrt{i}e^{-i\pi s^2}$$

and, separating the real and imaginary parts, that

$$\cos \pi x^2 \supset 2^{-1/2}(\cos \pi s^2 + \sin \pi s^2)$$
$$\sin \pi x^2 \supset 2^{-1/2}(\cos \pi s^2 - \sin \pi s^2).$$

5. (*a*) Apply the similarity and shift theorems. (*b*) $\{f\} = \{0\ 1\frac{1}{2}\ 3\ 4\frac{1}{2} \ldots 10\frac{1}{2}\ 12\ 11\frac{1}{2} \ldots 1\frac{1}{2}\ 1\ \frac{1}{2}\ 0\}$. (*c*) Agreement is better than 0.25 percent of $F(0)$ at the worst.

6. The agreement with $G(s) = 8e^{-i9\pi s}\operatorname{sinc} 8s + 4.5\delta(s)$ is not very good. Suppose that in a similar case you were unable to obtain the theoretical transform so easily but that you could compute it. In preparation for such a day, experiment with ways of recomputing the present case so that the computed values are closer to the theoretical transform.

7. The product has a limiting value of zero for all integer values of s up to N since there is one null factor for each such integer. For $s << 1$, the product approximates $1 - s^2 - s^2/4 - s^2/9 - \ldots = 1 - s^2(1 + 1/4 + 1/9 + \ldots) = 1 - (\pi^2/6)s^2$, which is also approximated by the Gaussian $\exp(-\pi^2 s^2/6)$. Consequently computation will show agreement with Gaussian only for s values so small that the quadratic expansion for the exponential function is good enough. Such poor agreement is to be expected from the standard expansion

$$\sin \pi s = \pi s \prod_{k=1}^{\infty}(1 - s^2/k^2),$$

which shows that the limit of the original continued product is sinc s, as mentioned in problem 8.32.

8. Assign the weekly mean $(1/7)\Sigma_j$ to F_0. The fundamental component, with a frequency of one cycle per week, will have both an amplitude and phase; assign these to F_1 and F_2. Assign F_3 and F_4 to the component with a frequency of two cycles per week, and the two numbers for the third harmonic, at three cycles per week, will complete the set of seven.

This raises an interesting question. The highest frequency discernible from daily values is 0.5 cycles per day, or 3.5 cycles per week, and the amplitude and phase of this two-day periodicity would have been ascertainable from the long data set that was used to compile g_j. However, the frequencies associated with the seven sufficient numbers F_k were at 0, 1/7, 2/7, and 3/7; the highest of these does not reach 0.5 cycles per day. Thus information on a conceivably significant characteristic (day-to-day alternation) has been lost. The loss mechanism is that data for alternate Mondays would be in antiphase, and so daily alternation is canceled out by the weekly summation. Such a variation would be discernible by summing over a fortnight to obtain a data set of 14 values. Now, posing the original question in terms of the new F_k, $k = 0$ to 13, and assigning F_0 to the mean $(1/14)\Sigma F_k$, what would be the remaining 13 values of F_k needed to fully represent the 14 data values?

CHAPTER 8: THE TWO DOMAINS

1. We may convolve the two Gaussians as in Problem 6.14 or use the following reasoning. Each of the Gaussians has area $\pi^{1/2}$; therefore their convolution will have area π because of the property

$$\int_{-\infty}^{\infty} (f * g)\, dx = \int_{-\infty}^{\infty} f\, dx \int_{-\infty}^{\infty} g\, dx. \tag{1}$$

(This is the property used for checking on convolution performed numerically (p. 32) and is derivable from the definite integral relation (p. 136), according to which both the left and right hand sides of (1) are equal to $F(0)\,G(0)$.) Now since variance is additive under convolution (p. 143) the variance of the desired convolution is double that of $\exp(-x^2)$; therefore it is $2^{1/2}$ times wider; multiplication by $\pi^{1/2}/2^{1/2}$ will bring the area to π as required. Hence

$$\exp(-x^2) * \exp(-x^2) \supset (\pi/2)^{1/2} \exp(-x^2/2).$$

This type of reasoning is valuable (*a*) for cross-checking results because it is rather unlike integration and therefore not likely to be subject to the same kind of mistakes and (*b*) for obtaining partial results, such as widths alone, when peak values or areas may be fixed by other considerations. In the above discussion we are shooting for *parameters;* we already know that Gaussians convolve into Gaussians.

6. The actual daily values are deducible and must be as follows, beginning January 1, 1900:

10, 10, 5, 0, 0, 0, 0, 0, 0, 0, 0, 15, 35, 30, 20, 25, 15, 10, 10, 10, 0, 0, 10, 10, 15, 25, 20, 20, 15, 20, 15.

In general, inversion of running totals is subject to an uncertain term consisting of a strictly periodic variation with a period equal to the length of the totalling period, and with zero mean value. Very often there are external facts helping to determine the amplitude and form of such an additive term. In the present problem the key fact is that the numbers represent sunspots and therefore cannot be negative. Consequently, on any occasion when the five-day running total is zero, the five adjacent daily values must be zero, and in January 1900 this enabled a *unique* solution to be found. On other occasions, our prior knowledge that there is no five-day periodicity in sunspots could be used. For example, by guessing a few consecutive values we could arrive at a solution, then numerically filter out frequencies around integral multiples of one-fifth of a cycle per day. This would lead to a fine optimization problem, since the limited quantity of data prevents the achievement of indefinitely narrow filter bands.

10. The definition proposed attempts to extend the applicability of the equivalent width in two ways. First, the integral of the spectrum $F(s)$ is replaced by the integral of the power spectrum $F(s)F^*(s)$ which will eliminate the cancellation that otherwise occurs if $F(s)$ is in antiphase with $F(0)$ at some values of s. Then, to allow for the possibility that the spectrum may peak up about a value other than $s = 0$, normalization is made to depend upon $F_{\max}$. However, the definition does not have the dimensions of bandwidth which should have the same units as s, and this will cause trouble. For exam-

ple, if a given $F(s)$ is amplified by a factor 2 at each value of s, which would not ordinarily be considered a band-width change, the quantity proposed increases by a factor 2.

13. (*a*) Let $g(x)$ be a shuffled form of $f(x)$. Then $\int g(x)\,dx = \int f(x)\,dx$ and $\int gg^* dx = \int ff^* dx$. Hence, referring to the definition,

$$W_{f\star f^*} = \int f\,dx \int f^* dx \Big/ \int ff^* dx = W_{g\star g^*}.$$

(*b*) The equivalent width W_g of $g(x)$ is, however, sensitive to $g(0)$. So

$$W_g = \int g\,dx/g(0) = \int f\,dx/g(0)$$

which is equal to W_f only if $f(0) = g(0)$.

(*c*) The total energy of a waveform is $\int\int ff^* dx$, which is unchanged by shuffling. Hence $|F(s)|^2$ at $s = 0$ is an invariant. Also $W_{|F|^2}$ is constant because $W_{|F|^2} = [W_{f\star f}]^{-1}$. Likewise $\int |F|^2\,ds$ is constant.

15. If $f(x) = e^{-x}H(x)$ then the symmetrized function is $g(x) = e^{-2|x|}$. Hence $e^{-|x|}$ is the same as the symmetrized function but with abscissas expanded by a factor 2. From this relationship we find that the power spectrum of $e^{-|x|}$ has twice the area (twice the waveform energy), half the equivalent width, and four times the central value.

18. By the modulation theorem

$$e^{-\pi x^2}\cos\omega x \supset \frac{1}{2}e^{-\pi(s+\omega/2\pi)^2} + \frac{1}{2}e^{-\pi(s-\omega/2\pi)^2}$$

This transform consists of two humps of unit equivalent width and if $\omega >> 1$, the humps will be widely separated. The transform of the desired autocorrelation function will be the square of the transform above and will also consist of two humps:

$$\frac{1}{4}e^{-2\pi(s+\omega/2\pi)^2} + \frac{1}{4}e^{-2\pi(s-\omega/2\pi)^2}$$

plus a cross-product term which is small. Inverting the transform we find the autocorrelation function to be

$$2^{-3/2}e^{-\pi x^2/2}\cos\omega x.$$

26. Consider a pressure variation

$$0.01\,e^{-\pi(t/T)^2}\cos(2500\,e^{t/5})\ \mathrm{N\,m^{-2}},$$

where T is say 10 s. It represents a tone that waxes and wanes in intensity and rises through one octave each 3.5 s. If this tone is repeated indefinitely at intervals of 3.5 s then at any given moment there will be several frequencies present going up by factors of 2. The impression will be of a complex note with harmonics, and as time elapses the pitch of the note will appear to rise. As each harmonic in turn rises out of audibility a subharmonic enters quietly from below, to gain in strength and to carry for a time the continuing impression of rising pitch.

29. Since $g(x) = f(x) * \Pi(x)$

$$g\left(x-\frac{1}{2}\right)+g\left(x-\frac{3}{2}\right)+g\left(x-\frac{5}{2}\right)+\ldots$$
$$= f(x) * \left[\Pi\left(x-\frac{1}{2}\right)+\Pi\left(x-\frac{3}{2}\right)+\Pi\left(x-\frac{5}{2}\right)+\ldots\right]$$
$$= f(x) * H(x).$$

Differentiating,

$$g'\left(x-\frac{1}{2}\right)+g'\left(x-\frac{3}{2}\right)+\ldots = f(x)$$
$$\left[\delta'\left(x-\frac{1}{2}\right)+\delta'\left(x-\frac{3}{2}\right)+\ldots\right] * g(x) = f(x).$$

Hence
$$\Pi^{-1}(x) = \delta'\left(x-\frac{1}{2}\right)+\delta'\left(x-\frac{3}{2}\right)+\delta'\left(x-\frac{5}{2}\right)+\ldots.$$

We can verify directly that

$$\Pi(x) * \left[\delta'\left(x-\frac{1}{2}\right)+\delta'\left(x-\frac{3}{2}\right)+\ldots\right] = \delta(x). \tag{1}$$

Since $(\operatorname{sinc} s)^{-1}G(s) = F(s)$ where $\operatorname{sinc} s \neq 0$, it might seem that $\Pi^{-1}(x) \supset (\operatorname{sinc} s)^{-1}$. However, we cannot divide by sinc s where sinc s = 0, so we cannot know $F(s)$ at $s = \pm 1, \pm 2, \ldots$ and for a good reason. Any Fourier component of $f(x)$ having an integral number of periods fitting into the width of the rectangle function is suppressed, is absent from the given function $g(x)$, and therefore cannot be restored. When $\Pi^{-1}(x) * g(x)$ has been calculated it is subject to a periodic additive term of unit period and arbitrary shape and magnitude signaling irreversibility, in much the same way as an arbitrary integration constant does.

This characteristic irreversibility shows up in the s-domain derivation at the point where division by zero alerts us. The x-domain argument conceals a subtle point. Rewriting (1) for a finite number of terms we get

$$\Pi(x) * \left[\delta'\left(x-\frac{1}{2}\right)+\delta'\left(x-\frac{3}{2}\right)+\ldots+\delta'\left(x-\frac{n}{2}\right)\right] = \delta(x) - \delta\left(x-\frac{n}{2}-\frac{1}{2}\right).$$

As $n \to \infty$ it would seem that the negative impulse would move out to the right beyond our ken, for all x. But, if we convolve both sides with some operand before letting $n \to \infty$ we may not get the same answer as convolving directly with $\delta(x)$. The exceptions will be sinusoids of period 1, 1/2, 1/3, ... which will be wiped out by the negative impulse.

In many cases the parameters of any periodic component can be determined from *a priori* information. For example, if a function $f(x)$ has stretches where it is zero, as at the beginning and end of finite-duration signals, any *periodic* component, which by definition never dies out, must have zero amplitude. The above discussion does not deal with restoration of deterministic signals in the presence of errors (*Proc. I.R.E.*, vol. 46, pp. 106–111, 1958), which in practice dominate what may be done.

32. We see that $(1 - x^2)(1 - x^2/4)(1 - x^2/9) = 1 - (1 + 1/4 + 1/9)x^2 + (1/4 + 1/9 + 1/36)x^4 - x^6/36 = 1 - 1.3611\,x^2 + 0.0274\,x^4 - 0.027\,x^6$, (1)
whereas

$$\operatorname{sinc} x = 1 - 1.6449\,x^2 + 0.8117\,x^4 - 0.1908\,x^6 + \ldots.$$

The coefficient of x^2 determines the central curvature (and the second moment of the Fourier transform), and the correct value is $-\pi^2/6$. In the infinite-product formula,

each successive factor increases the central curvature; consequently, no finite number of factors gives the correct coefficient. We may however note that the series Σ_n^{-2}, which appears as the coefficient of x^2 in (1), has a sum to infinity of $\pi^2/6$. In connection with the central limit theorem (p. 168) it was found that the product of many parabolic curves of the form $1 - ms^2$, not necessarily all with the same value of m, would approach a Gaussian function under certain conditions. Here we find a sinc function instead.

34. The desired variance is given by

$$\sigma^2 = \frac{\int_{-\infty}^{\infty} x^2 e^{-(\pi x/W)^2} \cos 2\pi\nu x\, dx}{\int_{-\infty}^{\infty} e^{-(\pi x/W)^2} \cos 2\pi\nu x\, dx}.$$

To avoid the integrations, work in terms of the Fourier transform $F(s)$ of the wavepacket $f(x)$. Then

$$\sigma^2 = -\frac{F''(0)}{4\pi^2 F(0)}.$$

From the modulation theorem

$$F(s) = \tfrac{1}{2}We^{-\pi[W(s+\nu)]^2} + \tfrac{1}{2}We^{-\pi[W(s-\nu)]^2}$$

and $F(0) = We^{-\pi W^2\nu^2}$. Note from a sketch of $F(s)$, which takes the form of two Gaussian humps of width $1/W$ centered at $s = \pm\,\nu$ (see Pictorial Dictionary), that the second derivative at the origin, $F''(0)$, is twice the second derivative of $\exp(-\pi W^2 s^2)$ at $s = \nu$. So

$$F''(0) = W(4\pi^2W^2s^2 - 2\pi)e^{-\pi W^2\nu^2}\big|_{s=\nu} = W(4\pi\nu^2 - 2\pi)e^{-\pi W^2\nu^2}.$$

Hence,

$$\sigma^2 = -\frac{W(4\pi^2W^2\nu^2 - 2\pi)e^{-\pi W^2\nu^2}}{4\pi^2We^{-\pi W^2\nu^2}}$$

$$= \frac{1}{2\pi} - W^2\nu^2.$$

If $W\nu > 1/\sqrt{2\pi}$ is σ^2 negative? Yes. When $W = \nu = 1$; then $\sigma^2 = -0.841$.

35. Begin by looking for a simple way to construct the function.

(*a*) Note that $g'(x)$ is expressible as a sum of four rectangle functions

$$\Pi[(x+b)/c] - \Pi[(x-b)/c] - (a^{-1} - 1)\Pi[(x + \tfrac{1}{2}a)/a],$$

where $b = (1 + a)/2$ and $c = 1 - a$. The transform of this, equal to $i2\pi sF(s)$, gives

$$F(s) = (c \operatorname{sinc} cs \sin 2\pi bs - (1 - a) \operatorname{sinc} as \sin \pi as)/\pi s.$$

(*b*) The second derivative $g''(x) = \delta(x + 4) + \delta(x - 4) - (a^{-1} - 1)[\delta(x + 1) + \delta(x - 1)] + 4(a^{-1} - 1)\delta(x)$. Therefore

$$F(s) = [2\cos 2\pi 4s - 2(a^{-1} - 1)\cos 2\pi s + 4(a^{-1} - 1)]/(-4\pi^2 s^2).$$

(*c*) Subtracting $(1 - a)\Lambda(x/a)$ from the trapezoid $(1 - a)\Pi[(x/(1 + a)] * \Pi[x/(1 - a)]/(1 - a)$ offers another construction, from which

$$F(s) = (1 - a)(1 + a)\operatorname{sinc}[(1 + a)s] \times \operatorname{sinc}[(1 - a)s] - a(1 - a)\operatorname{sinc}^2 as.$$

(*d*) Noting that $f(x)$ is the difference of two triangle functions, $f(x) = \Lambda(x) - \Lambda(x/a)$, gives

$$F(s) = \operatorname{sinc}^2 s - a \operatorname{sinc}^2 as.$$

This last is obviously the preferred answer. It is striking how different the various versions look, even though they are all equal. But they are not equal as regards the amount of work required to derive and verify. The moral is to begin by spending time imagining different approaches.

36. (*a*) The length would be such that the round-trip travel time on the line is $\Delta = 0.1\ \mu s$, i.e., 15 m. (*b*) The input impedance Z to a transmission line of length L, open-circuited at the far end, is

$$Z = Z_0 \coth\left[\sqrt{(r + i\omega l)(g + i\omega c)}L\right],$$

and the usual approximation for a loss-free line ($r = g = 0$) in terms of the inductance l and capacitance c per unit length is

$$Z = -iZ_0 \cot 2\omega T,$$

where $T = 2vL$ and $v = 1/\sqrt{lc}$ is the wave velocity on the line (close to the speed of light). The corresponding voltage transfer function is $-i \cot 2\omega T$. (*c*) The impulse response is $I(t) = \delta(t) + 2\delta(t - \Delta) + 2\delta(t - 2\Delta) + \ldots.$

The transfer function of an open-circuited length of transmission line is the Fourier transform of $I(t)$. Since $I(t)$ is not an odd function of t, the transfer function cannot be pure imaginary but must have a real part. This seems to imply that a loss-free circuit can dissipate power, but in reality an electric circuit always dissipates some power. (See *Electronics Letters*, vol. 34, pp. 1927–1928, 1998.)

37. In Fig. 8.16 the second curve can be written in terms of $E(x) = \exp(-x)H(x)$ as $-[E(x + 5) - E(x - 5)]$, which is a finite difference.

The first curve cannot be expressed in this way as a finite difference of a single hump taken over a long interval. However, the curve possesses a finite sum in terms of which it is therefore expressible as a finite difference. A finite sum $s(x)$ of a function $f(x)$ is $f(x) * \{1\,1\,1\,1\ldots\} = f(x) + f(x - a) + f(x - 2a) + f(x - 3a) + \ldots.$ The finite difference of $s(x)$ over the same interval a, yields the original $f(x)$. Finite summing and finite differencing are thus reciprocal operations, analogous to integration and differentiation. Sequential application of the two operators $\{1\,1\,1\,1\ldots\}$ and $\{1\,1\} *$ yields the identity operator. Thus $\{1\,1\,1\,1\ldots\} * \{1\ {-1}\} = \{1\}$.

The third example can be expressed as the derivative of the function with four steps that is the integral of the given set of four impulses, but cannot be expressed as a *finite* difference.

CHAPTER 9: WAVEFORMS, SPECTRA, FILTERS, AND LINEARITY

11. (*a*) Consider the history of the input for the whole day, and the corresponding output. If we began again tomorrow with an identical box and the same day's worth of input the whole output would be the same, delayed by one day. Therefore the system is time-invariant. Jones did not take account of the fact that the state of the system was different in the afternoon, possibly because someone else used it at lunchtime and applied a step voltage.

(*b*) The system is nonlinear. Understand that the gate switch has to be open before the second input $V_2(t)$ is applied, so that all tests are performed on the same initial state of the system, and opened again, if necessary, before applying the sum $V_1(t) + V_2(t)$ of the two separate inputs whose superposition is to be tested.

12. The system is time-invariant, by the definition of time-invariance. We can see that superposition breaks down because the compressive force is proportional to the charge on each plate and therefore to the square of the voltage. Doubling the voltage will not therefore merely double the response, but will produce something different. Therefore the system is nonlinear.

26. The relation follows from the derivative theorem in reverse, namely $-i2\pi x f(x) \supset F'(s)$.

27. (*a*) The function $y(x)$ is strictly periodic, with period 2π, or frequency $\frac{1}{2}\pi$, as can be seen by eye from the plot. (*b*) The Fourier transform $Y(s)$ has four spikes, at $s = \pm\frac{5}{2}\pi$ and $\pm\frac{4}{2}\pi$. This follows from addition of the transforms $4.5\delta(s \pm \frac{5}{2}\pi)$ and $5.5\delta(s \pm \frac{4}{2}\pi)$. The transform $Y(s)$ does not contain a component at the frequency $\frac{1}{2}\pi$. We conclude that periodicities may be apparent to the eye that are not detectable by Fourier analysis. This phenomenon is particularly striking in two- and three-dimensional crystal structures.

28. The transform of the sum of two functions equals the sum of the transforms, and therefore no linear combination of the two functions can contain any frequency that is not present in either of the original functions. The present problem, however, is nonlinear because the two component functions $y_1(x)$ and $y_2(x)$ do not simply add; the final printing is given by $y(x) = H[y_1(x) + y_2(x)]$. Where one black line overprints another the result is black; the resulting value of $y(x)$ is 1, not 2. In this case, Fourier analysis will reveal the visible periodicity.

29. (*a*) $L_{crit} = W/\sqrt{2\pi} = 0.79788W$. The minimum disappears where the second derivative of $\exp(-\pi x^2/W^2)$ is zero.

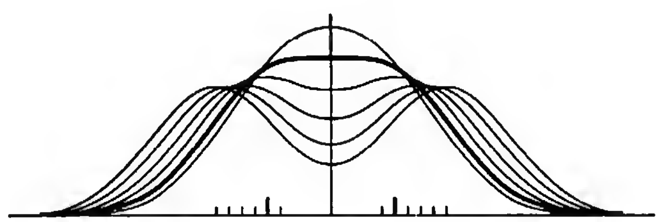


Fig. 9.29 **Pairs of equal Gaussians, of equivalent width W and symmetrically placed at $\pm\frac{1}{2}L$ as indicated by ticks, are summed. The central curvature of the composite profile changes sign when $L = L_{crit} = 0.8W$ (heavy ticks and curve). The spectrograph record for lines spaced by L_{crit} shows a flat top with no central minimum. The adjacent curves, above and below the flat topped curve, are for $0.8L_{crit}$ and $1.2L_{crit}$ respectively, while the outlying curve is for $2L_{crit}$.**

(*b*) The single peak resulting when two "unresolved" lines coalesce is not Gaussian in shape; the separation and amplitudes of the two components can be extracted in

principle. But in practice, where the observed profile is the sole source of information, who would assure the observer that the "true" distribution consisted of just two equal-strength lines? To know that is to have foreknowledge of the unknown. As an alternative, the observer might postulate the existence of a small number of lines with arbitrary positions and strengths and show that an unequal line pair was compatible with observation within the errors, and use the conclusion in the design of a further observation. In the presence of instrumental or external errors, one could establish the resolvable separation of an equal line pair as a function of error level, assuming plausible error statistics. No doubt it would be hard to improve much over L_{crit}.

Finally, another given of the original problem as worded was the assertion that the instrumental profile was known. That is sometimes effectively the case in practice where an instrument can be tested with a single strong narrow line, but even then the instrumental response is fundamentally a measured profile subject to its own errors; could one ever assume that an instrumental response was Gaussian out to more than two or three equivalent widths from line center?

30. (*a*) $V(0) = 0$
$V(1) = 0.5$
$V(2) = e^{-1} + 0.8$
$V(3) = e^{-2} + 1.6e^{-1} + 1$
$V(4) = e^{-3} + 1.6e^{-2} + 2e^{-1} + 0.3$
$V(5) = e^{-4} + 1.6e^{-3} + 2e^{-2} + 0.6e^{-1}$
$V(6) = e^{-5} + 1.6e^{-4} + 2e^{-3} + 0.6e^{-2}$

(*b*) The coefficients are exactly the same as the values of $V_2(t)$ at integer values of time, as expected from the fact that digital signal theory is included within continuous-time theory as supplemented by delta notation.

CHAPTER 10: SAMPLING AND SERIES

6. The integral of $|sF(s)|$ over a limited band may differ seriously from the infinite integral, because of the factor s, even though the fraction of power μ beyond the band limits is small. The corresponding view in the x-domain is that a small amplitude sinusoid, added to a band-limited function, can seriously affect the net slope if the frequency of the sinusoid is high.

17. (*a*) A filter containing only a finite number of inductors, capacitors and resistors has a finite number of natural modes and its natural behavior is therefore specified by a finite number of constants.

(*b*) It will be evident when sufficient coefficients for use in $Y_i = \alpha_1 Y_{i-1} + \alpha_2 Y_{i-2} + \cdots + \beta X_i$ have been determined because the right hand side will then correctly predict the output Y_i.

35.
```
function[y]=binom(n)
b=([1 1]); y=1; count=1;
while count < n+1
   y=conv(y,b); count=count+1;
end
```

36. Yes. So we see that Mountararat's lofty desire for simplification has suppressed the factor $1/L$ in one place only to have it rebound in two places in the transform domain. The transform of the string of unit impulses spaced L is $\mathrm{III}(Ls)$, as obtainable directly from the similarity theorem. It seems a retrograde step to adopt notation that makes a fundamental theorem inapplicable, while at the same time not saving any ink.

37. The sample values of the spectrum $S(f)$ suffice to obtain other values by interpolation. For example, if the samples straddle $f = 0$ symmetrically, one can get $S(0)$ from

$$S(0) = 0.63S\left(\frac{1}{2}T^{-1}\right) - 0.21S\left(1\frac{1}{2}T^{-1}\right) + \ldots + 0.63S\left(-\frac{1}{2}T^{-1}\right) - 0.21S\left(-1\frac{1}{2}T^{-1}\right) + \ldots.$$

The coefficients applied to the samples in this case are those for midpoint interpolation.

38. The sum of the series must be even and also periodic with period 2π. Outside the range $0 \le x \le \pi$, the sum of the series must be $\frac{1}{2}\pi|\sin x|$. The fundamental period is thus π. The coefficients are found by integrating $\frac{1}{2}\pi \sin x \cos 2\pi n x$ from 0 to 2π.

39. The summation has its poles coinciding with those of $\cot x$, and, if the proposed identity is correct, the zeros would fall at $\pm\pi/2, \pm 3\pi/2, \pm 5\pi/2, \ldots$. Examine $\sum_{-\infty}^{\infty}(x - k\pi)^{-1}$ to see if it is zero at $x = \pi/2$. The series to be checked is

$$\sum_{-\infty}^{\infty} = \frac{1}{\pi/2 - k\pi} = \frac{2}{\pi}\sum\frac{1}{1-2k}$$
$$= \frac{2}{\pi}\left[1 + \frac{1}{-1} + \frac{1}{-3} + \frac{1}{-5}\ldots + \frac{1}{3} + \frac{1}{5} + \frac{1}{7} + \ldots\right]$$

If the terms are selectively paired the sum can be made zero, but can be made to have other values too; for this reason the series cannot be claimed to converge.

The selective pairing of $(x - k\pi)^{-1}$ with $(x + k\pi)^{-1}$ can be made explicit by summing each pair to obtain $2x/(x^2 - k^2\pi^2)$. The resulting formula

$$\cot x = \frac{1}{x} + 2x\sum_{k=1}^{\infty}\frac{1}{x^2 - k^2\pi^2}$$

is then correct for all x except those values where $\cot x$ has a pole. The impeccability of this expansion results from forcing the sequence of summing.

One expects problems arising from the fact that $\cot x$ has an infinite number of infinite discontinuities. If these are removed by convolution with $\epsilon^{-1}\exp(-\pi x^2/\epsilon^2)$, where ϵ is a very small positive number, the resulting smoothed function can be made as indistinguishable from $\cot x$ as we please. Consequently the proposed identity has some virtue; thus, it can be used with the convolution theorem to produce such correct results as

$$\epsilon^{-1}e^{-\pi x^2/\epsilon^2} * \cot x \supset e^{-\pi\epsilon^2 s^2}\mathrm{III}(s)\operatorname{sgn} s.$$

The trouble is this. Convolving x^{-1} with a string of unit impulses spaced by π entails adding samples of a function x^{-1} that has area $\pm\infty$ under its tails. If equal and opposite samples are neutralized by preaddition then a finite sum for the convolution results; but the teaching of our mathematical forbears is that, with infinite sums, bundling the terms can lead to discordant results. There is no unique sum.

40. The graph will show, if it was not otherwise obvious, that $\Sigma_k \sin k$ does not converge, but will also show that the partial sums oscillate evenly about the value $\frac{1}{2}\cot\frac{1}{2}^{\circ}$.

Replication of $1/x$ at intervals π, the operation supposed to add up to cot x, may be expressed by $x^{-1} * \pi^{-1}\mathrm{III}(x/\pi)$, whose FT is $(-i\pi \operatorname{sgn} s)\mathrm{III}(\pi s)$, which back transforms into $\Sigma_1^\infty 2 \sin 2kx$. Thus a consequence of the supposed representation of cot x by superposed $1/x$ functions would be that $\cot x = \Sigma_1^\infty 2 \sin kx$.

As in the preceding problem, incorporation of the suggested transform as a factor in an expression that is physically more realistic can lead to a rigorously correct result.

CHAPTER 11: THE DISCRETE FOURIER TRANSFORM AND THE FFT

Solutions to Exercises *a* to *e* appearing in the body of Chapter 11 are presented first and are followed by solutions to the numbered problems.

(*a*) We find that **fft** ([4 2 1 0 0 0 1 2]) yields

$$\{10 \quad 6.8284 \quad 2 \quad 1.1716 \quad 2 \quad 1.1716 \quad 2 \quad 6.8284\},$$

which is purely real, as stated.

(*b*) The eight samples of $\Lambda(t/4)$ for $t = -3, -2, \ldots, 3, 4$ are

$$\begin{array}{ccccccc} \{1 & 2 & 3 & 4 & 3 & 2 & 1\} \\ \scriptstyle -2 & \scriptstyle -1 & \scriptstyle 0 & \scriptstyle 1 & \scriptstyle 2 & \scriptstyle 3 & \scriptstyle 4 \\ \scriptstyle 6 & \scriptstyle 7 & \scriptstyle 8 & & & & \end{array}$$

where the small numbers immediately below the samples are indices n that are allowed to go negative, just as t does. Apply the rule for replacing $-n$ by $8 - n$ to obtain the positive indices 6, 7, 8 in the second row of small numbers. The representation of the even function $\Lambda(t/3.5)$ using positive indices is then seen to be {4 3 2 1 0 1 2 3}. In general, for an N-element sequence, the replacement rule is

$$x(-n) \Rightarrow x(N - n),$$

for those values of n such that $1 \leq N - n \leq N$.

(*c*) Consider an even function $f(t)$ for which $f(0) = 0, f(\pm 1) = 1, f(\pm 2) = 2, f(\pm 3) = 3$ and $f(\pm 4) = 4$. By the replacement rule, the corresponding 9-element sequence, indexed starting with $n = 1$, is $\{f(1) \ f(2) \ f(3) \ f(4) \ f(-2) \ f(-1) \ f(0)\} =$ {1 2 3 4 4 3 2 1 0}. More than one function of a continuous variable possesses this set of samples. The simplest example in this case is $f(t) = 4.5\{\Pi(t - \frac{1}{2})/9] - \Lambda(t/4.5)\}$ whose Fourier transform is $40.5 \operatorname{sinc} 9s - 20.25 \operatorname{sinc}^2 4.5s$. As a further exercise one could plot the real and imaginary parts of this Fourier transform and compare with what **fft( )** gives for the sample sequence.

(*d*) The given sequence represents a sample set of $\Lambda(t/4) \operatorname{sgn} t$ at unit intervals of t. Applying **fft( )** produces $\{0 \quad -9.6569i \quad -4i \quad -1.6569i \quad 0 \quad 1.6569i \quad 4 \quad 9.6569i\}$, which is purely imaginary as expected. The presence of the additional term $\sin \pi t$ would not change the sample values at the integer locations.

(*e*) The three discrete transforms are

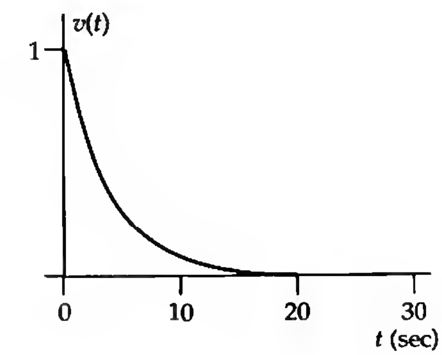
$X_4(k) = \{4\ \ 0\ \ 0\ \ 0\}.$

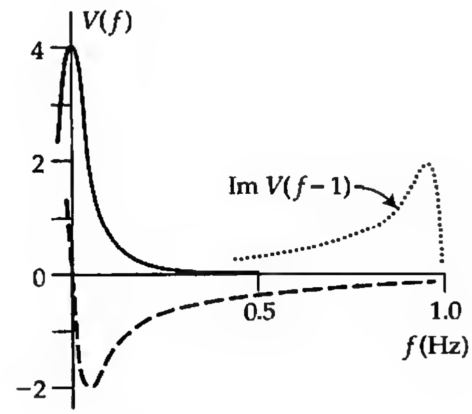
$X_8(k) = \{4\ \ \ 1 - 2.4142i\ \ \ 0\ \ \ 1 - 0.4142i\ \ \ 0\ \ \ 1 + 0.4142i\ \ \ 0\ \ \ 1 + 2.4142i\}.$

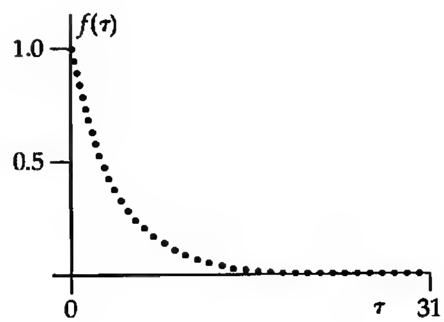
$X_{16}(k) = \{4\ \ \ 3.0137 - 2.0137i\ \ \ 1 - 2.4142i\ \ \ -0.2483 - 1.2483i\ \ \ 0\ \ \ 0.8341 + 0.1659i$
$1 - 0.4142i\ \ \ 0.4005 - 0.5995i\ \ \ 0\ \ \ 0.4005 + 0.5995i\ \ \ 1 + 0.4142i$
$0.8341 - 0.1659i\ \ \ 0\ \ \ -0.2483 + 1.2483i\ \ \ 1 + 2.4142i\ \ \ 3.0137 + 2.0137i\}.$

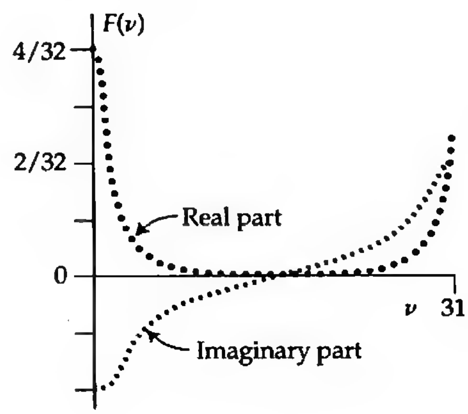
For comparison, $\mathrm{Re}F(s) = 4 \cos 5\pi \operatorname{sinc} 4s$ and $\mathrm{Im}F(s) = -4 \sin 5\pi s \operatorname{sinc} 4s$.

10. The transform $V(f) = 4(1 + i8\pi f)^{-1}$ is illustrated by its real and imaginary parts. It will be noticed that the imaginary part dies out much less rapidly than the real part. (That is because the odd part of $v(t)$ is discontinuous.) The real and imaginary parts of $F(\nu)$ and the DFT are also shown. The two real parts agree very well in the range $0 \leq f \leq 0.5$ corresponding to $0 \leq \tau \leq 16$. However, in the vicinity of $f = 0.5$ the agreement is not good as regards the imaginary parts.









This can be understood in terms of aliasing that would arise in the imaginary part of $V(f)$ if one were to take samples at one-second intervals. Because Im $V(f)$ has not died out by $f = 0.5$ Hz, one-second sampling would be inadequate. Inclusion of the replicated island Im $V(f - 1)$, shown dotted, accounts for the main discrepancy between FT and DFT around $f = 0.5$, in this case. In general, more distant islands $V(f - n)$ may contribute noticeably also.

13. (*a*) The discrete Hartley M-file is

```
function y=dht (f,N)
% Discrete Hartley transform of N-element real vector f
F=fft(f);
y=(real(F)-imag(F))/N;
```

(*b*) Applying the command **dht** twice in succession produces the original vector as expected except that it is eight times too small. To remedy this, omit the division by N. For the inverse operation **idht** the M-file is

```
function y=idht(f,N)
% Inverse discrete Hartley transform of N-element real vector f
F=fft(f);
y=(real(F)-imag(F));
```

14. (*a*) Taking the central element 5 to correspond to $x = 0$, we see that the given elements are values of $5(1 - |x|/5)$ for $x = 0, \pm 1, \pm 2, \ldots$. But the nine elements do not constitute the full infinite range of samples. (*b*) The index ν corresponds to frequency ν/N in cycles per unit of x, where in this example $N = 20$ elements. Thus the first null at $\nu = 4$ corresponds to frequency 4/20, in agreement with that expected from $s = 1/5$. Agreement between DFT and FT is perhaps reasonable; but at $\nu = 6$, $s = 0.03$, where the maximum of the first sidelobe of sinc^2 occurs with a value $25 \times 0.045032 = 1.1258$, the DFT gives 1.5279. This is because of the overlap of repeats such as the one seen peaking at $\nu = 20$. These overlaps can be pushed further away by insertion of more zero samples.

ν	0	1	2	3	4	5	6	...	20
20 × DFT	25	20.43	10.47	2.43	0	1.00	1.53	...	20.43
s	0	1/20	2/20	3/20	4/20	5/20	6/20	...	1
$25\,\text{sinc}^2 5s$	25	20.26	10.13	2.25	0	0.81	1.13	...	20.43

15. The sequence {1 4 6 4 1} matches $f(x) = (16/\sqrt{2\pi})\exp(-\frac{1}{2}x^2)$ as to both r.m.s. width and area. The transform $F(s) = 16\exp(-2\pi s^2)$ gives values in rough agreement with the DFT as follows:

ν	0	1	2	3	...	8	...	15	16
16 × DFT	16	14.80	11.66	7.65	...	0	...	14.80	16
s	0	1/16	2/16	3/16	...	8/16	...	15/16	1
$16\exp(-2\pi^2 s^2)$	16	14.81	11.75	7.99	...	0.115	...	$5/10^7$	$4/10^8$

One sees that if the FT is to be obtained by taking the DFT of samples then accuracy is a major concern. Values of the DFT can be compared graphically with $6 + 8\cos 2\pi s + 2\cos 2\pi 2s$, the transform of $\delta(x+2) + 4\delta(x+1) + 6\delta(x) + 4\delta(x-1) + \delta(x-2)$, to see the effects of replication.

16. The ordinary convolution sum is

$$\{1\ 12\ 65\ 208\ 429\ 572\ 429\ 0\ -429\ -572\ -429\ -208\ -65\ -12\ -1\}$$

with the cyclic convolution sum is

$$\{-64\ 0\ 64\ 208\ 429\ 572\ 429\ 0\ -429\ -572\ -429\ -208\}.$$

18. (*a*) A MATLAB program is

```
f = [1 4 6 4 1];
F = fft(f);
G = zeros(1,10);
G(1,3) = F(1:3);
G(9:10) = F(4:5);
h = ifft(G)*2
```

(*b*) Multiply the result by 2 to compensate for the change from 10 values to 5.
(*c*) Change line 3 to `G = zeros(1:30)` and change line 5 to `G(29:30) = F(4:5)`
(*d*) The MATLAB interpolates, for comparison, are

{.1273 **0** .2970 **0** .5516 **1** 2.4190 **4** 5.4323 **6** 5.4323 **4** 2.4190 **1** .5516 **0** .2970 **0** .1273}

CHAPTER 12: THE HARTLEY TRANSFORM

10. (*a*) {0 0 0 0 −2 0 0 0}
There is no d.c. value; the only frequency present is 0.5; consequently the only content of the DHT is at $\nu/N = 0.5$, or $\nu = 4$.
(*b*) The power spectrum is {0 0 0 0 4 0 0 0}. Since there is no content at $\nu = 1$ and 7, 2 and 6, 3 and 5, all the power spectrum values are zero except at $\nu = 4$.

11. The Fourier series $r = r_0 + \Sigma a_n \cos n\theta$ can represent a square as closely as wished if enough terms are used. The cas function formula is not only a simple exact equation for a square but it is in a directly computable form. Using θ mod $\pm\frac{1}{4}\pi$ is also interesting.

12. In the case of the Fourier transform, $\mathcal{F}\mathcal{F}f(x) = f(-x)$ and the function is reversed, so four transformations are required to return to the original $f(x)$. The cyclicity of the Hartley transform is already known from the reciprocal property of this transform.

20. Although the infinite integral of $\delta'(t)$ is zero, the integral from zero to infinity is not, nor is the integral from minus infinity to infinity of the absolute value. Consequently the derivative of a delta function is not a null function. Nor would we want it to be. If you hit a golf ball of mass m with a force $A\delta'(t)$ you would give a velocity $(A/m)\delta(t)$ and produce a displacement (A/m) for $t > 0$, whereas if you hit something massive with a null function it does nothing.

21. The impulse is located to the ENE. The period of the cas function transform is $(a^2 + b^2)^{-\frac{1}{2}} = 2d$. Thus $a = \cos 22.5°/2d$ and $b = \sin 22.5°/2d$.

22. Let $F_{\text{real}} = x$ and $F_{\text{imag}} = y$. In the coordinate system (x',y') which is rotated through θ relative to (x,y),

$$x' = x \cos\theta + y \sin\theta$$
$$y' = y \cos\theta - x \sin\theta.$$

When $\theta = -\pi/4$,

$$\sqrt{2}x' = x - y = F_{\text{real}} - F_{\text{imag}} = H(\nu)$$
$$\sqrt{2}y' = y + x = F_{\text{imag}} + F_{\text{real}} = H(-\nu).$$

Thus $H(\nu) + iH(-\nu)$ is derivable from $F_{\text{real}} + iF_{\text{imag}}$ by rotating through $-\pi/4$ and magnification by $\sqrt{2}$.

23. From the formula for the tangent of a sum we have

$$\begin{aligned}\tan(\phi + \pi/4) &= [\tan(\pi/4) + \tan\phi]/[1 - \tan(\pi/4)\tan\phi] \\ &= (1 + \tan\phi)/(1 - \tan\phi) \\ &= (1 + F_{\text{imag}}/F_{\text{real}})/(1 - F_{\text{imag}}/F_{\text{real}}) \\ &= (F_{\text{real}} + F_{\text{imag}})/(F_{\text{real}} - F_{\text{imag}}) \\ &= H(-\nu)/H(\nu).\end{aligned}$$

24. Put $a = \cos\theta, b = -\sin\theta, c = 0, d = \sin\theta, e = \cos\theta, f = 0$.

26. The value of $^{\text{II}}F(0)$ should equal the mean of the data values divided by $\sqrt{2}$. For the inverse DCT2 there is nothing comparable in simplicity. The relation is $f(0) = 2^{-1/2}\,{}^{\text{II}}F(0) + {}^{\text{II}}F(1)\cos(\pi/2N) + {}^{\text{II}}F(2)\cos(2\pi/2N) + {}^{\text{II}}F(3)\cos(3\pi/2N) + \dots$.
For DCT1 the leading value is $N^{-1/2}[f(0)/\sqrt{2} + f(1) + f(2) + \dots + f(N-1) + f(N)/\sqrt{2}]$, and, since DCT1 is symmetrical, the value of $f(0)$ can be checked similarly.

CHAPTER 13: RELATIVES OF THE FOURIER TRANSFORM

42. Convert to the Fourier transform of the even function $f(|x|)$ and evaluate $\int_{-\infty}^{\infty} f(|x|)\cos 2\pi sx\,dx$. Since $f(|x|) = \frac{1}{2}\Pi(x) + \frac{1}{2}(1 - 4x^2)\Pi(x)$ we can use the known transform of a rectangular function plus a parabolic pulse to find

$$F_c(s) = \frac{1}{2}\operatorname{sinc} s + \frac{1}{\pi^2 s^2}(\operatorname{sinc} s - \cos \pi s).$$

Check by noting that $F_c(0) = 5/6$, which equals the area under $f(|x|)$.

43. Since you do not know the purpose of the processing it might be effective, in getting the job, to list questions that reveal your capabilities. For example, is computing speed desired? Something done once a month does not sound urgent, but there may be a publishing deadline shortly after close of business at the end of the month. So, if speed is of the essence, you will develop a fast algorithm based on factorization of 365 as 5×73. Are daily fluctuations in $f(\tau)$ to be deemphasized? You note that the monthly mean has been suppressed; explain what you would do about the 7-day component that is caused by weekends. Is the choice of the cosine function intended to suppress the odd part of $f(\tau)$? If there is a monthly cycle (such as accounting practice introduces), the end-of-month trend will be suppressed. You will investigate the effect of varying the origin of τ. Finally say how you would invert the transformation, which was the ostensible assignment, trying to say something that will give you an advantage over other prospective consultants whose reports will be compared with yours.

44. A ray traversing an inhomogeneous medium with a slowly varying refractive index n suffers curvature equal to the gradient of n in the direction transverse to the ray. For rays only lightly inclined to the z-axis, $d^2r/dz^2 \approx dn/dr = -(2n_0/h^2)r$. Recalling that $d^2r/dx^2 = -\omega^2 r$ has solution $A\cos\omega x + B\sin\omega x$, applying the boundary conditions $r = 0$ and $dr/dz = \tan i$ at $z = 0$ gives

$$r = \omega^{-1}\tan i\,\sin\omega x,$$

where $\omega = \sqrt{2n_0}/h$ and is related to the quarter period L by $\omega = \pi/2L$. Hence $L = \pi h/2\sqrt{2n_0}$.

CHAPTER 14: THE LAPLACE TRANSFORM

3. (*a*)

$$\begin{aligned} F(p) &= \int_{-\infty}^{\infty} \Pi(t)e^{-pt}\,dt = \int_{-1/2}^{1/2} e^{-pt}\,dt \\ &= -p^{-1}[e^{-pt}]_{-1/2}^{1/2} = -p^{-1}[e^{-p/2} - e^{p/2}] \\ &= 2p^{-1}\sinh(p/2),\ \text{all } p. \end{aligned}$$

(*b*)

$$F(p) = \int_0^1 e^{-pt}\,dt = -p^{-1}(e^{-p} - 1),\ \text{all } p.$$

(*c*)

$$\begin{aligned} F(p) &= \int_{-\infty}^{\infty} \Lambda(t)e^{-pt}\,dt = \int_{-1}^{1} (1 - |t|)e^{-pt}\,dt \\ &= \int_{-1}^{0} (1 + t)e^{-pt}\,dt + \int_{0}^{1} (1 - t)e^{-pt}\,dt = \ldots . \end{aligned}$$

These simple integrals may be evaluated in several more lines but application of the convolution theorem (p. 385) to (*a*) yields $F(p) = 4p^{-2}\sinh^2(p/2)$, all p.

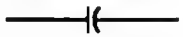
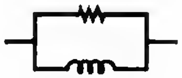
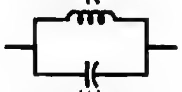
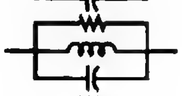
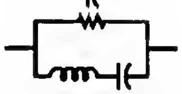
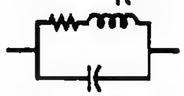
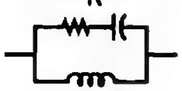
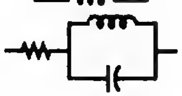
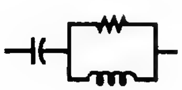
5. See p. 540.

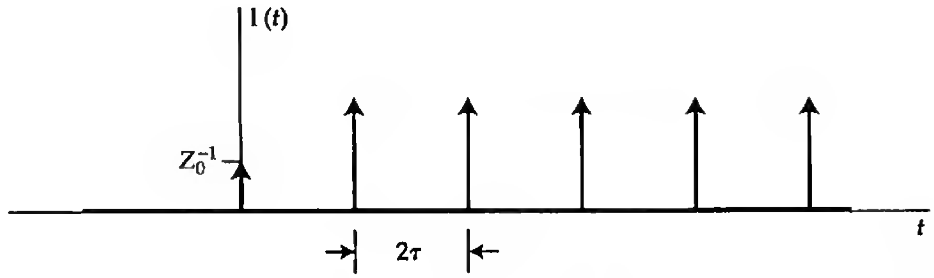
8. Using the relation $\tan ix = i\tanh x$ and remembering that the given impedance is reactive we may write the cable impedance as $Z(p) = 75\tanh(10^{-8}p/2\pi) = Z_0\tanh\tau p = Z_0(1 - e^{-2\tau p})/(1 + e^{-2\tau p})$ where $\tau = 10^{-8}/2\pi$ seconds $= 0.6$ nanoseconds is the time taken for an electromagnetic wave to travel the length of the cable. Evidently the cable is a convenient piece about 12 cm long.

$$\begin{aligned} I(t) &= [Z(p)]^{-1}V(t) = [Z(p)]^{-1}\delta(t) \\ &= \frac{Z_0^{-1}}{1 - e^{-2\tau p}}\delta(t) + e^{-2\tau p}\frac{Z_0^{-1}}{1 - e^{-2\tau p}}\delta(t) \\ &= \frac{Z_0^{-1}}{2\tau}\mathrm{III}\left(\frac{t}{2\tau}\right)H\left(\frac{t}{2\tau} + \frac{1}{2}\right) + \frac{Z_0^{-1}}{2\tau}\mathrm{III}\left(\frac{t - 2\tau}{2\tau}\right)H\left(\frac{t - 2\tau}{2\tau} + \frac{1}{2}\right). \end{aligned}$$

The first term comes from Table 14.2. The second term, which is an identical train of impulses delayed by a time 2s, the round-trip echo time, comes from the shift theorem. The net result is as shown:

Solution to Problem 5.

Circuit	Admittance	Impulse Response	Step Response
	R^{-1}	$R^{-1}\delta(t)$	$R^{-1}H(t)$
	$L^{-1}p^{-1}$	$L^{-1}H(t)$	$L^{-1}tH(t)$
	Cp	$C\delta'(t)$	$C\delta(t)$
	$L^{-1}(p + R/L)^{-1}$	$L^{-1}e^{-Rt/L}H(t)$	$L^{-1}(1 - e^{-Rt/L})H(t)$
	$R^{-1}p(p + 1/RC)^{-1}$	$R^{-1}\delta(t) - L^{-1}e^{-Rt/L}H(t)$	$R^{-1}H(t) - R^{-1}(1 - e^{-Rt/L})H(t)$
	$L^{-1}p(p^2 + 1/LC)^{-1}$	$L^{-1}\cos[(LC)^{-1/2}t]H(t)$	$(C/L)^{1/2}\sin[(LC)^{-1/2}t]H(t)$
	$R^{-1} + L^{-1}p^{-1}$	$R^{-1}\delta(t) + L^{-1}H(t)$	$R^{-1}H(t) + L^{-1}tH(t)$
	$R^{-1} + Cp$	$R^{-1}\delta(t) + C\delta'(t)$	$R^{-1}H(t) + C\delta(t)$
	$L^{-1}p^{-1} + Cp$	$L^{-1}H(t) + C\delta'(t)$	$L^{-1}tH(t) + C\delta(t)$
	$R^{-1} + L^{-1}p^{-1} + Cp$	$R^{-1}\delta(t) + L^{-1}H(t) + C\delta'(t)$	$R^{-1}H(t) + L^{-1}tH(t) + C\delta(t)$
	$R^{-1} + L^{-1}p(p^2 + 1/LC)^{-1}$	$R^{-1}\delta(t) + L^{-1}\cos[(LC)^{1/2}t]H(t)$	$R^{-1}H(t) + (C/L)^{1/2}\sin[(LC)^{-1/2}t]H(t)$
	$Cp + L^{-1}(p + R/L)^{-1}$	$C\delta'(t) + L^{-1}e^{-Rt/L}H(t)$	$C\delta(t) + R^{-1}(1 - e^{-Rt/L})H(t)$
	$L^{-1}p^{-1} + R^{-1}p(p + 1/RC)^{-1}$	$L^{-1}H(t) + R^{-1}\delta(t) - L^{-1}e^{-Rt/L}H(t)$	$L^{-1}tH(t) + R^{-1}H(t) - R^{-1}(1 - e^{-Rt/L})H(t)$
	$R^{-1} - R^{-2}C^{-1}p(p^2 + p/RC + 1/LC)^{-1}$	$R^{-1}\delta(t) - R^{-2}C^{-1}(a - b)^{-1}(ae^{-at} - be^{-bt})H(t)$ $a,b = (2RC)^{-1} \mp [(2RC)^{-2} - (LC)^{-1}]^{1/2}$	$R^{-1}H(t) + R^{-2}C^{-1}(a - b)^{-1}(e^{-bt} - e^{-at})H(t)$
	or (equal roots), $4R^2 = L/C$	$R^{-1}\delta(t) - R^{-2}C^{-1}(1 - ct)e^{-ct}H(t)$, $c = (2RC)^{-1}$	$R^{-1}H(t) - R^{-2}C^{-1}te^{-ct}H(t)$
	$L^{-1}(p + 1/RC)(p^2 + p/RC + 1/LC)^{-1}$	$L^{-1}(b - a)^{-1}[(c - a)e^{-at} - (c - b)e^{-bt}]H(t)$, a, b as above, $c = 1/RC$	$L^{-1}(b - a)^{-1}[-a^{-1}(c - a)(1 - e^{-at})$ $+ b^{-1}(c - b)(1 - e^{-bt})]H(t)$
	or (equal roots), $4R^2 = L/C$	$L^{-1}[1 + (c - d)t]e^{-dt}H(t)$, $c = 1/RC$, $d = (2RC)^{-1}$	$L^{-1}[cd^{-2}(1 - e^{-dt}) - (c - d)d^{-1}te^{-dt}]H(t)$
	$R^{-1} + (R^2C - L)(R^2LC)^{-1}[p - R(R^2C - L)^{-1}]$	$(p^2 + p/RC + 1/LC)^{-1}$ $R^{-1}\delta(t) + (R^2C - L)(R^2LC)^{-1}(b - a)^{-1}[(c - a)e^{-at}$ $- (c - b)e^{-bt}]H(t)$, a,b as above, $c = R(L - R^2C)^{-1}$	$R^{-1}H(t) + (R^2C - L)(R^2LC)^{-1}(b - a)^{-1}$ $[-a^{-1}(c - a)(1 - e^{-at})$ $+ b^{-1}(c - b)(1 - e^{-bt})]H(t)$
	or (equal roots), $4R^2 = L/C$	$R^{-1}\delta(t) + (R^2C - L)(R^2LC)^{-1}[1 + (c - d)t]e^{-dt}H(t)$, $c = -R(R^2C - L)^{-1}$, $d = (2RC)^{-1}$	$R^{-1}H(t) + (R^2C - L)(R^2LC)^{-1}[cd^{-2}(1 - e^{-dt})$ $- (c - d)d^{-1}te^{-dt}]H(t)$
	$L^{-1}p(p^2 + pR/L + 1/LC)$	$(2L)^{-1}(a^2 - b^2)^{-1/2}\{[a + (a^2 - b^2)^{1/2}]$ $\exp[-a - (a^2 - b^2)^{1/2}t] - [a - (a^2 - b^2)^{1/2}]$ $\exp[-a + (a^2 - b^2)^{1/2}t]\}H(t)$ $2a = R/L$, $b^2 = 1/LC$	$(2L)^{-1}(a^2 - b^2)^{-1/2}\{-\exp[-a - (a^2 - b^2)^{1/2}t]$ $+ \exp[-a + (a^2 - b^2)^{1/2}t]\}H(t)$



The doubling that occurs after the first current pulse is due to momentary superposition of a reflected wave on the echo.

Only the operational layout for this problem is given as the conventional presentation is too bulky.

10. By inspection $I(p) = \dfrac{1}{R + 1/Cp} \cdot \dfrac{p}{p^2 + \omega^2}$

$$= R^{-1} \frac{p^2}{(p + \alpha)(p^2 + \omega^2)} \qquad \text{where } \alpha = 1/RC$$

$$= R^{-1}\left[\frac{A}{p + \alpha} + \frac{B}{p + i\omega} + \frac{C}{p - i\omega}\right]$$

where $A = \alpha^2/(\alpha^2 + \omega^2)$, $B = \omega/2i(\alpha - i\omega)$ and $C = -\omega/2i(\alpha + i\omega)$. Hence

$$I(t) = R^{-1}[\alpha^2(\alpha^2 + \omega^2)^{-1}e^{-\alpha t} + Be^{-i\omega t} + Ce^{i\omega t}]\,H(t)$$
$$= R^{-1}(\alpha^2 + \omega^2)^{-1}[\alpha^2 e^{-\alpha t} + \omega^2 \cos \omega t - \omega\alpha \sin \omega t]\,H(t)$$
$$= (1 + \omega^2C^2R^2)^{-1}[R^{-1}e^{-\alpha t} + R\omega^2C^2 \cos \omega t - \omega C \sin \omega t]\,H(t).$$

13. Suppose that steady direct current $I(t) = H(t)$ appears at $t = 0$ in a parallel combination of L and C for which $Z(p) = Lp(LCp^2 + 1)^{-1}$ and calculate the voltage $V(t)$ from the operational expression

$$V(t) = Z(p)\,H(t)$$
$$= Z(p)\,p^{-1}\delta(t)$$
$$= C^{-1}(p^2 + 1/LC)^{-1}\,\delta(t).$$

From the table of transforms we see that

$$V(t) = C^{-1}\omega^{-1} \sin \omega t\, H(t).$$

14. Consider a series L, C circuit for which $Z(p) = Lp + (Cp)^{-1}$ through which a current $\sin \omega t\, H(t)$ flows. The voltage $V(t)$ will be expressible in operational notation as

$$V(t) = [Lp + (Cp)^{-1}] \sin \omega t\, H(t).$$

Now from Table 14.2

$$\sin \omega t\, H(t) = \omega(p^2 + \omega^2)^{-1}\,\delta(t).$$

Hence

$$\begin{aligned} V(t) &= [Lp + (Cp)^{-1}]\,\omega(p^2+\omega^2)^{-1}\,\delta(t) \\ &= \frac{LCp^2+1}{Cp(p^2+\omega^2)}\omega\delta(t) \\ &= L\omega p^{-1}\,\delta(t) \quad \text{if} \quad \omega^2 = 1/LC \\ &= L\omega H(t). \end{aligned}$$

Thus, the flow of a.c. in this circuit is associated with the appearance of a d.c. voltage across it. The direct voltage exists entirely across the capacitance. Alternating voltages appear across the L and C but there is no net alternating voltage across the combination.

21. The standard explanation of existence of several time functions with the same Laplace transform refers to the inversion integral (p. 381) and points out that the path of integration defined by the constant c may have several significantly different relationships to the poles of the transform. Thus in the problem at hand where poles occur at $p = 2$, 1 and -1 we may choose $2 < c$, $1 < c < 2$, $-1 < c < 1$ or $c < -1$ and obtain different results on carrying out the integration.

An alternative explanation starts with the theorem, that if for real α

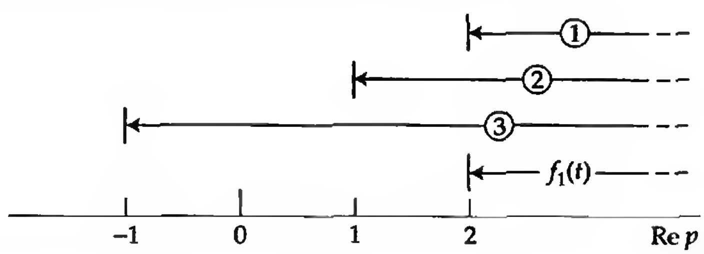
$$e^{-\alpha t}H(t) \supset F(p) \qquad -\alpha < \operatorname{Re} p$$

then

$$-e^{-\alpha t}H(-t) \supset F(p) \qquad \operatorname{Re} p < -\alpha.$$

Now $f_1(t) = \left[\frac{1}{3}e^{2t} - \frac{1}{2}e^{t} + \frac{1}{6}e^{-t}\right]H(t) \supset \dfrac{1}{(p-2)(p-1)(p+1)} \qquad 2 < \operatorname{Re} p.$

The region of convergence is set by the term in e^{2t} as we may note from the diagram showing the regions of convergence of all three terms



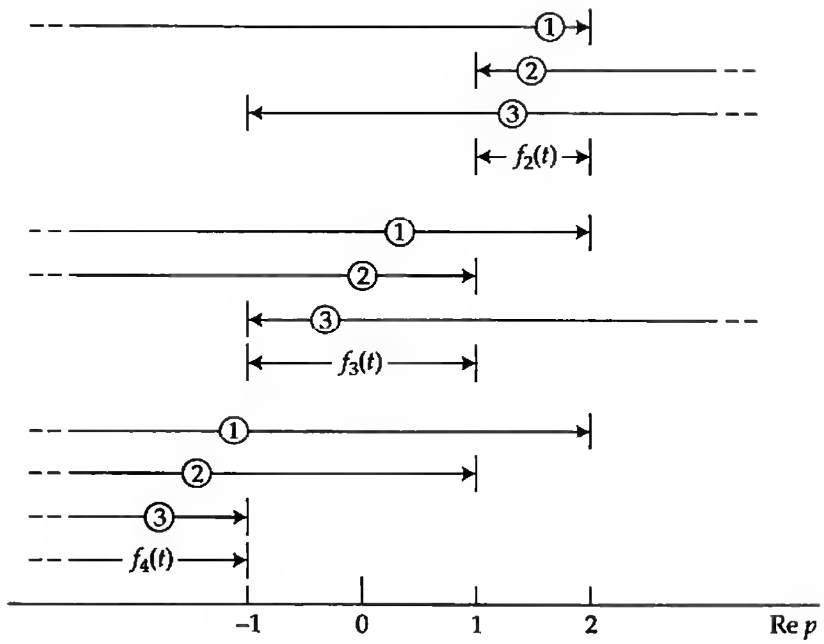
Now consider the following functions derived from $f_1(t)$ by altering one or more terms as shown:

$$f_2(t) = -\frac{1}{3}e^{2t}H(-t) - \frac{1}{2}e^{t}H(t) + \frac{1}{6}e^{-t}H(t) \qquad 1 < \operatorname{Re} p < 2$$

$$f_3(t) = -\frac{1}{3}e^{2t}H(-t) + \frac{1}{2}e^{t}H(-t) + \frac{1}{6}e^{-t}H(t) \qquad -1 < \operatorname{Re} p < 1$$

$$f_4(t) = -\frac{1}{3}e^{2t}H(-t) + \frac{1}{2}e^{t}H(-t) - \frac{1}{6}e^{-t}H(-t) \qquad \operatorname{Re} p < -1.$$

The following diagrams show the existence of strips of convergence common to all three terms.



23. If there was no stored energy before the voltage $v(t)H(t)$ was applied, all integrals from $0-$ to ∞ would be the same as integrals from $-\infty$ to ∞ because all integrands from $-\infty$ to $0-$ would be zero. Then the situation would be the same as for the two-sided Laplace transform and the theorems for the special transform would be the same as in Table 14.1 for the two-sided transform. There would be no occasion, as in the example set out on p. 396, to include terms representing initial behavior $f(0+), f'(0+)$, etc.

But if energy is present prior to the application of $v(t)H(t)$ an adjustment will be needed. The values of $f(0-), f'(0-), \ldots$ will suffice to define the prior state. The special transform of $f'(t)$ may be evaluated by integration by parts as follows.

$$\int_{0-}^{\infty} e^{-pt} f'(t)dt = [e^{-pt}f(t)]_{0-}^{\infty} + p\int_{0-}^{\infty} e^{-pt} f(t)dt$$
$$= -f(0-) + pF_-(p).$$

Likewise

$$\int_{0-}^{\infty} e^{-pt} f''(t)dt = [e^{-pt}f'(t)]_{0-}^{\infty} + p\int_{0-}^{\infty} e^{-pt} f'(t)dt$$
$$= -f'(0-) - pf(0-) + p^2F_-(p).$$

Example. A voltage source in series with a resistance R and capacitance C generates a voltage $V_0H(t)$ but just before the voltage jumps from zero to V_0 a current I_0 is flowing. Find the subsequent current.

The differential equation is

$$C^{-1}I(t) + RI'(t) = V_0\delta(t).$$

The subsidiary equation is

$$C^{-1}\bar{I}(p) + R[p\bar{I}(p) - I_0] = V_0.$$

The bars refer to the special transform, and we have used the derivative theorem. Solving for $\bar{I}(p)$ we have

$$\bar{I}(p) = \frac{V_0 + RI_0}{Rp + C^{-1}}$$

$$= (V_0/R + I_0)\frac{1}{p + R^{-1}C^{-1}}.$$

Inverting the transform we have

$$I(t) = (V_0/R + I_0)e^{-t/RC} \qquad t > 0.$$

(Verify that the transform of $e^{-t/RC}$ is the same whether the lower limit of integration is 0− or 0+.)

Comment. Note that post-initial values such as $I(0+)$ were not needed. It seems more natural to be given the prior state and the excitation and to be asked to find what happens than to be given the excitation and *part* of what then happens, with the task of finding out the rest of what happens.

25. Transforming the differential equation term by term

$$sY(s) + Y(s) = 0$$

$$(s + 1)Y(s) = 0$$

$$Y(s) = 0 + k\delta(s + 1)$$

because $Y(s)$ can contain a delta function of any strength k at $s + 1 = 0$. Inverting the transform,

$$y(t) = ke^{-t}$$

and the initial conduction fixes the value of k at $y(0+)$.

26. This interesting problem illustrates two philosophical points. First let us approach it by forming the subsidiary equation

$${}_0\bar{I}(p)Lp + {}_0\bar{I}(p)R = {}_0\bar{V}(p).$$

The zero subscript indicates one-sided Laplace transform. No term in $I(0+)$ is required because at $t = 0$ the circuit was as yet quiescent. Thus the first comment is that the appearance of initial values on the RHS in the traditional layout of initial value problems (p. 396) is restricted to situations where the switching on takes place at $t = 0$. In a virtually identical problem where switching on occurs later, these terms are dispensed with. And yet we expect to get the solution nonetheless. Continuing, we solve for ${}_0\bar{I}(p)$ to find

$${}_0\bar{I}(p) = \frac{{}_0\bar{V}(p)}{Lp + R}.$$

Let us take a particular case where $V(t) = H(t - 1)$ and we know the solution must turn out to be

$$I(t) = R^{-1}(1 - e^{-R(t-1)/L})H(t - 1).$$

Since

$$H(t - 1) \supset p^{-1}e^{-p}$$

$$
\begin{aligned}
{}_0\bar{I}(p) &= \frac{p^{-1}e^{-p}}{Lp + R} \\
&= R^{-1}e^{-p}\frac{R/L}{p(p + R/L)}.
\end{aligned}
$$

From p. 388 this gives

$$I(t) = R^{-1}(1 - e^{-(R/L)(t-1)})H(t - 1),$$

which is the correct response.

Had the voltage been applied at $t = 0$, it would have been necessary to give $I(0+)$ in order to get a solution; but in the present problem it was not necessary to be given the analogous $I(1+)$.

The second comment is, why should it be necessary to be given partial information about the response current, such as initial values, when the response is what we are charged with discovering. Does it ever happen in practice when we need to find out what current will flow that a circuit (or equivalently its differential equation) is given, that the exciting voltage is given, and that *in addition* some intelligent agent is able to furnish $I(0+)$, $I'(0+)$, $I''(0+)$, . . .? In other words, in some way, someone knew what was about to happen next. Possibly the reason for the appearance of such unreal problems in books is that traditional boundary value exercises in standard texts on differential equations (such as finding the trajectory of a ball when height and tilt of the cannon are given) have been unthinkingly translated from the space to the time domain.

27. Let $I(t) = \delta(t) + a\delta(t - 1) + a^2\delta(t - 2) + \dots$.

Then

$$
\begin{aligned}
\bar{I}(p) &= 1 + ae^{-p} + a^2e^{-2p} + \dots \\
&= (1 - ae^{-p})^{-1}.
\end{aligned}
$$

In terms of frequency, the transfer function $T(f)$ is given by

$$T(f) = (1 - ae^{-i2\pi f})^{-1}.$$

This result could be produced by a resistance in parallel with a lossy, nondispersive, short-circuited transmission line.

30. (*a*) Express $\mathrm{III}(t)H(t)$ as the sum of its even and odd parts: $\mathrm{III}(t)H(t) = \frac{1}{2}\mathrm{III}(t) + \frac{1}{2}\operatorname{sgn} t$. The transform of $\frac{1}{2}\mathrm{III}(t)$ is $\frac{1}{2}\mathrm{III}f$, which supplies the missing even part of the transform.

■

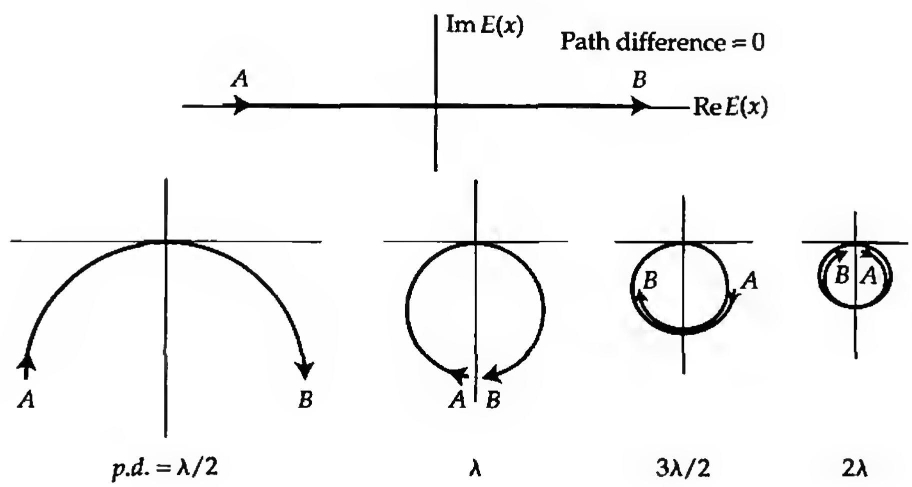
CHAPTER 15: ANTENNAS AND OPTICS

1. Two antennas with beamwidths that are equal by some customary measure such as the angle between half-power directions may have quite different equivalent widths of their angular spectra. For example, since $P(s)$ can in some directions be opposite in sign to $P(0)$, the value of $\int P(s)\,ds$ can be sensitive to the sign of sidelobes which have little to do with the concept of beamwidth. As an extreme example, the two aperture

distributions $\Pi(x/1000\lambda)$ and $\Pi(x/1000\lambda) - \Pi(x/10\lambda)$ are quite similar and have similar beams, but the equivalent width of the angular spectrum of the second is zero.

As a different kind of objection, an antenna, whose beam was shifted (by phasing) with no change of beamwidth at all (as measured in units of s between half-peak points), would exhibit a change in the equivalent width of $P(s)$ because $P(0)$ changed.

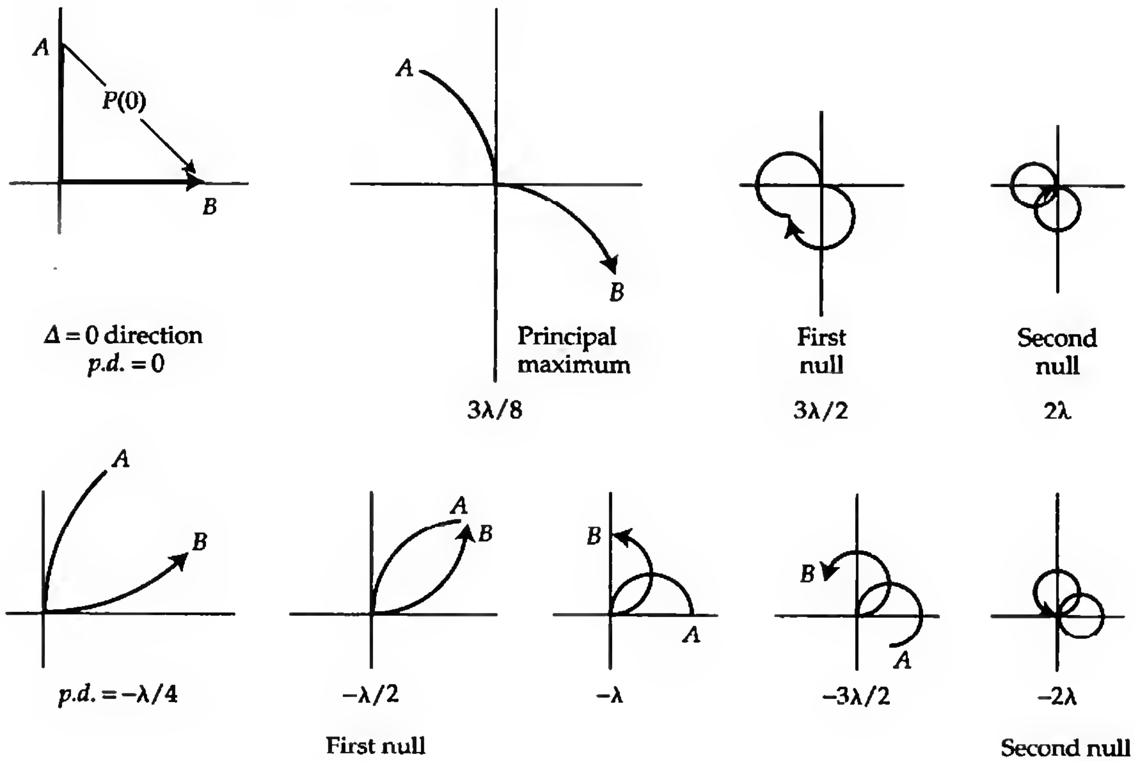
2. The diagrams are of the kind shown in Fig. 2.10 (p. 20) and in the answers to Problems 2.20 and 2.21. The Cornu spiral arising in optical diffraction is a diagram of this kind. Although the function domain and transform domain are normally regarded as distinct, it is interesting to note that the plane of these diagrams may be regarded equally as the complex plane of the aperture field E or of the angular spectrum P. The phasor AB is the resultant of the elementary phasors $E(x/\lambda)\,d(x/\lambda)\exp(-i2\pi sx/\lambda)$ and gives the angular spectrum in amplitude and phase for particular values of direction s. In each diagram the arc length is the same,



5. The first null on one side is in the direction where the path difference to the end elements of the aperture is $-\lambda/2$, and the second null is where the path difference is -2λ. The third null is at -2.5λ. In the direction normal to the aperture (path difference to ends = 0) the resultant AB has a magnitude $2^{-1/2}$ times the arclength AB. Moving away from the normal we can see graphically that the resultant decreases toward the side considered above (see p.d. = $-\lambda/4$) but increases toward the other side. Representing the aperture distribution by $E(\xi) = -i\Pi(2\xi - \frac{1}{2}) + \Pi(2\xi + \frac{1}{2})$, where $\xi = x/w$ and w is the full aperture width, we obtain for the angular spectrum

$$P(s) = (w/\lambda)\left[-ie^{-i\pi\sigma}\frac{1}{2}\operatorname{sinc}\sigma + e^{i\pi\sigma}\frac{1}{2}\operatorname{sinc}\sigma\right] \tag{1}$$

where $\sigma = ws/2\lambda$. To find the location of the principal maximum we can solve $dP/ds = 0$, which is rather tedious. Alternatively, we can establish by rough sketches that there is a maximum in the vicinity of p.d. = $3\lambda/8$. The magnitude may then be calculated to be 1.85 $P(0)$, and of course the phase is $-45°$ by symmetry.



7. By direct integration we obtain values for $n = 0$ to 7. The crude approximation described under central-limit theorem (p. 169) leads to an approximation $(n\pi/2)^{1/2}$ for large n

n	Effective beamwidth	$(n\pi/2)^{1/2}$	n	Effective beamwidth	$(n\pi/2)^{1/2}$
0	1		4	$8/3 = 2.67$	2.51
1	$\pi/2 = 1.57$	1.25	5	$15\pi/16 = 2.95$	2.80
2	2	1.77	6	$16/5 = 3.20$	3.06
3	$3\pi/4 = 2.36$	2.17	7	$35\pi/32 = 3.44$	3.32

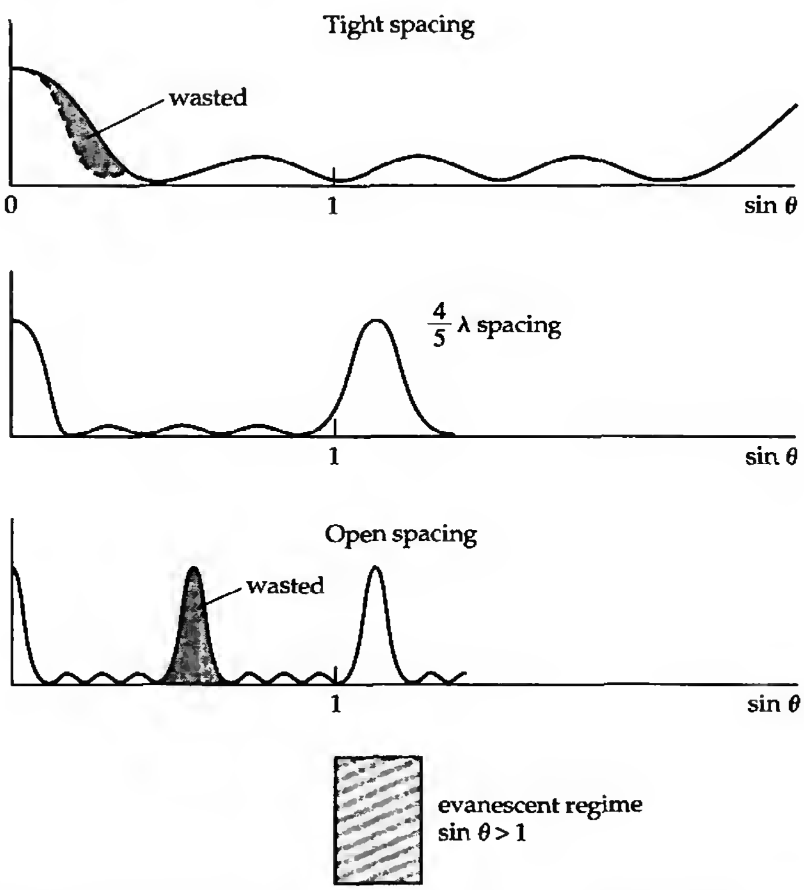
10. Because an antenna aperture has finite extent (say N wavelengths), its angular spectrum is deficient in spatial frequencies (measured in cycles per unit of s) exceeding $\frac{1}{2}N$, and its power radiation pattern contains no spatial frequencies higher than N. Consequently the radio image formed in the focal plane of a paraboloidal reflector of focal length f is correspondingly devoid of fine detail. The critical spatial sinusoid has a period N^{-1} radians on the sky and a period on the focal plane of fN^{-1} meters. The critical sampling interval is $\frac{1}{2}fN^{-1}$ meters. The collector normally found at the focus of a paraboloid is designed so that its beam fills the solid angle subtended at the focus by the rim of the paraboloid. The angle of the cone is $2\arctan(\frac{1}{2}N\lambda/f)$, or roughly $N\lambda/f$ radians, a beamwidth obtainable with a focal collector about fN^{-1} meters across. These calculations verify that the size of a focal collector could in fact be thought of as a device that is correctly sized to fit one image cell in the focal plane.

11. One way is to construct a pinhole camera, i.e., an opaque sheet of material containing a pinhole behind which is a photographic plate or other array of detectors.

If the intensity is so low as to require an unreasonably low exposure time the hole size may be increased. If the blurring due to increased hole size becomes the limiting consideration, increased sensitivity can be gained by the use of several small holes distributed somewhat at random. This will produce a superposition of faint images that can be combined, for example by autocorrelating the total image. As with all autocorrelation functions, there will be a sacrifice of phase information.

Another approach is to make use of the fact that x-rays do reflect at highly grazing incidence, which in effect lengthens the wavelength. A paraboloidal annulus remote from both the vertex and the focus will act as a lens.

16. It is true that the field intensity on the ground vertically below the aircraft is independent of the spacing of the slots (assuming that the slot fields are kept the same). However, it is wrong to deduce from this that it is permissible to pack the slots tightly, because the gain, which is to be a maximum, depends not only on the field intensity vertically below, but also on the input power required to maintain it. The structural expert, although possessing a wide background, does not take this into account and therefore his statement should be accorded only the respect due to an expert speaking outside his field.



The antenna expert's assistant, a mathematically inclined fellow, is right in saying that the power radiated is the sum of the powers passing through each slot, but of

course means time-averaged powers. Rayleigh's theorem, however, applies instant by instant, and the power flowing through the slots at one instant may flow back in a quarter cycle later. This would be the case with an aperture excited by a field that produces evanescent waves but does not radiate at all. The actual antenna problem involves some power flow of this kind.

To maximize the gain while keeping the field vertically below the same, we have to minimize

$$\int\int 2\pi P_n(\theta,\phi)\, d\Omega.$$

In the illustration, all cases show the same radiation on axis. The tight-spacing case shows a fat beam which means power is wasted in unwanted directions. The open-spacing case shows power being wasted in a grating lobe. The optimum occurs just as the first grating lobe is sliding over from the evanescent regime ($\sin\theta > 1$) to the radiating regime. This optimum is a flat one which in practice is of the order of four-fifths wavelength spacing.

18. Consider a ring of 96 equispaced two-dimensional unit delta functions on a radius a. The diametrically opposite pair at $(a, 2\pi p/96)$ and $(a, \pi + 2\pi p/96)$, where $p = 1, 2, \ldots, 48$, possess an angular spectrum in the form of a cosinusoidal corrugation with isolines perpendicular to the line joining the two delta functions, or $2\cos(2\pi a Q/\lambda)$ where Q is measured perpendicular to the isolines. For any point in the (l,m)-plane, Q is given by $Q = l\cos(2\pi p/96) + m\sin(2\pi p/96)$. Adding the contributions from the 48 pairs of deltas, we see that the angular spectrum of the 96 deltas is

$$\sum_{p=0}^{47} 2\cos\{(2\pi a/\lambda)[l\cos(2\pi p/96) + m\sin(2\pi p/96)]\}. \tag{1}$$

The summation, if carried to $p = 95$, would give double the result. Note that $a/\lambda = 214$ (not 430).

Now use q for radial variable in the (l,m) or angular spectrum plane, instead of r as in the problem wording, and reserve r for radial variable in the antenna plane, or (x,y)-plane.

The ensemble of point deltas, when the points are extremely numerous, is like a ring delta of strength 96 given by $96\,(2\pi a/\lambda)^{-1}\,\delta(r - a/\lambda)$ where r is the dimensionless radial coordinate, measured in wavelengths, in the antenna plane; in fact this ring delta can be thought of as deriving from the leading constant in a Fourier series $96\,(2\pi a/\lambda)^{-1}\,\delta(r - a/\lambda)(1 + \sum_1^{96} 2\cos n\theta)$ for the periodic function of angle constituted by the 96 deltas. This leading term transforms into a J_0 function as we know from the transform pair

$$\delta(r - a/\lambda)^2 \supset (2\pi a/\lambda)\, J_0(2\pi a q/\lambda) \qquad \text{(p. 338)}.$$

Later we establish the more general transform pair

$$\delta(r - a/\lambda)\cos 96n\theta^2 \supset (2\pi a/\lambda)\, J_{96n}(2\pi a q/\lambda)\cos 96n\theta,$$

where $r = \lambda^{-1}(x^2 + y^2)^{1/2}$ if x and y are rectangular coordinates in the antenna plane. Thus the angular spectrum is proportional to

$$J_0(2\pi a q/\lambda) + \sum_{n=1}^{96} 2\, J_{96n}(2\pi a q/\lambda)\cos 96\, n\theta.$$

If we are not dealing with delta functions, but elements whose angular spectrum is $P(q)$, then the foregoing "array factor" would be multiplied by $P(q)$ (p. 413).

The Bessel function $2J_{96}(2\pi aq/\lambda)$ has a maximum of 0.29 near $2\pi aq/\lambda = 99.72$ and is extremely small for smaller arguments, varying roughly as $(2\pi aq/\lambda)^{96}$. Its first zero is at $2\pi aq/\lambda = 104.72$. In this vicinity the second term J_{198} is absolutely negligible. In the vicinity of the origin both are negligible and so the diffraction pattern is proportional to J_0^2 to high precision. In the vicinity of $2\pi aq/\lambda = 100$, the envelope of $J_0(z)$, calculated from the asymptotic expression $(2/\pi z)^{1/2}$, is about 0.08 and therefore interacts with $2J_{96}(z)$ to produce important intensity variations.

The figure shows positive crest lines of the cosinusoidal corrugations corresponding to $p = 0, 1, 2$. Clearly the sum of such corrugations, near the origin, should approximate a J_0 function because J_0 can be defined as the limit of such a sum as the number of elements approaches infinity (see footnote p. 336). The next point of interest is at A where many of the positive crests reinforce. Exactly the same happens at B, C, etc. Similar reinforcement occurs again at 96 further points A', B', etc., points which are about twice the distance from the origin and so on to higher orders.

Midway between A and B troughs of the corrugations reinforce, but this does not lead to an intensity maximum in this case because corrugations at other values of p tend to have compensating crests there. Even so, (1) is a perfectly good formula for computing the intensity distribution. To see analytically what is happening in the first order grating response it is more direct to consider $[J_0(z) + 2J_{96}(z)\cos 96\,\theta]^2$. It is helpful to know that $J_0(z) \sim (2/\pi z)^{1/2}\cos(z - \pi/4)^{96}$, that the first maximum of $J_n(z)$ is near $z = n + 0.81n^{1/3}$ and the adjacent zero near $z = n + 1.86n^{1/3}$ (Abramovitz and Stegun, p. 371).

Appendix to 15.18

To show that

$$\delta(r - a/\lambda)\cos n\theta \;{}^2\!\supset\; (2\pi a/\lambda)\cos n\phi(-i)^n J_n(2\pi aq/\lambda)$$

we need to draw on the basic integral

$$J_n(z) = \frac{i^{-n}}{2\pi}\int_{-\pi}^{\pi}\cos n\beta\, e^{iz\cos\beta}\,d\beta$$

(Abramovitz and Stegun, p. 360, no. 9.1.21.)

$$\begin{aligned}
F(l,m) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,e^{-i2\pi(lx+my)}\,dx\,dy \\
&= \int_0^{\infty}\int_{-\pi}^{\pi} f(x,y)e^{-i2\pi qr\cos(\theta-\phi)}r\,dr\,d\theta \qquad \text{(p. 247)} \\
&= \int_0^{2\pi}d\theta\int_0^{\infty}\delta(r - a/\lambda)\cos n\theta\, e^{-i2\pi qr\cos(\theta-\phi)}r\,dr \\
&= \int_0^{2\pi}d\theta\cos n\theta\, e^{-i2\pi q(a/\lambda)\cos(\theta-\phi)}(a/\lambda).
\end{aligned}$$

Put $\beta = \theta - \phi$; then

$$F(l,m) = (a/\lambda)\int_0^{2\pi}\cos[n(\beta+\phi)]\,e^{-i2\pi q(a/\lambda)\cos\beta}\,d\beta$$

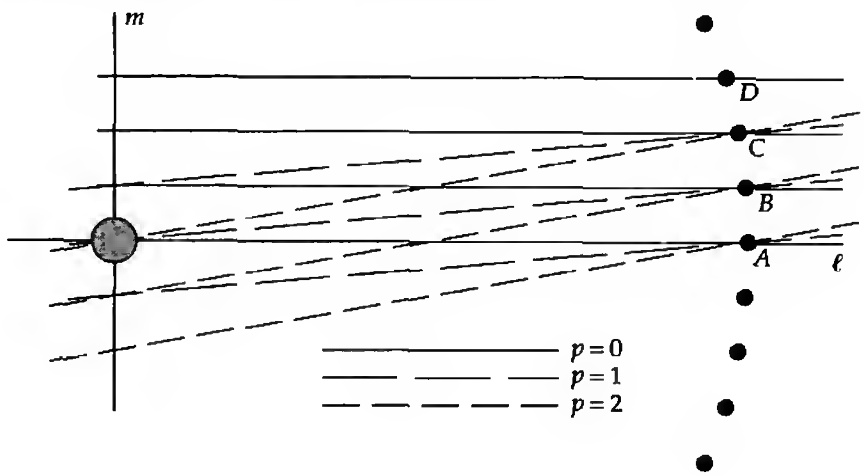
$$= (a/\lambda)\cos n\phi \int_0^{2\pi} \cos n\beta\, e^{-i2\pi q(a/\lambda)\cos\beta} d\beta + \text{null term}$$

$$= (2\pi a/\lambda)\cos n\phi\, i^n J_n(-2\pi aq/\lambda).$$

As a special case

$$\delta(r - a/\lambda)\cos\theta \;{}^2\!\supset\; -i(2\pi a/\lambda)\cos\phi\, J_1(2\pi aq/\lambda).$$

Rotating $f(x,y)$ in its own plane rotates the two-dimensional Fourier transform through the same angle. Therefore cos may be replaced on both sides by sin. By the addition theorem $\cos\theta$ may be replaced by $\exp i\theta$ and $\cos\phi$ by $\exp i\phi$.



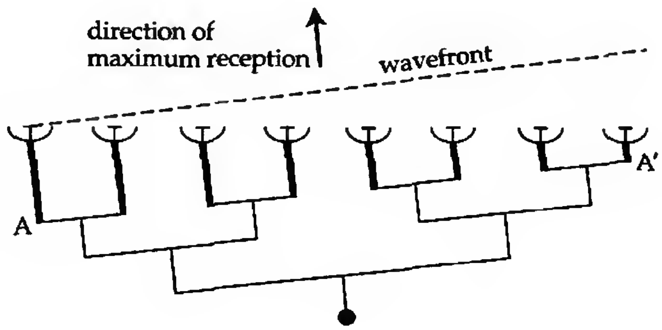
19. An antenna array may be correctly phased to point in a particular direction by the insertion of extra lengths of transmission line arranged as shown. If the transmission lines are air-filled, so that the wave velocity in the line is the same as in free space, then the extra lengths (shown heavy) are such that the transmission line layout has the direction AA' parallel to the wavefronts coming from the direction of maximum reception. Such an array can point in only one direction at a time because the phase gradient along the array is fixed by the transmission line lengths.

An array of pressure gauges on the sea bed could be connected together in a similar way by tubing; but if independent pressure records are made, they may be combined additively with artificial insertion of time delays equal to what would have been inserted by extra lengths of tubing. Thus an array of recording systems permits retrospective processing for several directions (eight?). With one receiver one looks in only one direction at a time.

The radiofrequency analogy using multiple receivers would require local-oscillator frequency stability to a faction of one radiofrequency period over the observing period, which is available at some expense. A more practical procedure is to divide the antenna outputs several ways and to apply the numerous receivers to the numerous two-antenna combinations that can be formed by direct interconnection (*Proc. IEEE*, vol. 61, no. 9, pp. 1249–1257, September 1973).

Direction finding on traveling ocean waves is feasible in principle in the open ocean but near the coast is subject to complications introduced by reflection and refraction. Wave velocity c depends on water depth h as given by $v^2 = (g\lambda/2\pi)\tan(2\pi h/\lambda)$. On the deep ocean, 1 km waves travel with a velocity of 39 ms^{-1}

(frequency of 0.39 Hz); but over the continental shelf (h = 200 m) the velocity drops and the wave direction is changed by refraction. In shallower coastal waters the wave direction will be nearly perpendicular to the coastline, independent of the original direction. Coastal scattering and reflection confuse the situation further.



21. Let e_j be the instantaneous voltage that the element labeled j delivers to the common point. The product of two radiofrequency voltages contains a steady or slowly-varying component of interest and a component at twice the radio frequency whose rejection will be indicated by time-averaging brackets $\langle\ \rangle$. The interferometer response is proportional to the time average of the product of the voltages delivered from the two sides, or

$$\langle(e_1 + e_2 + \ldots + e_{16})(e_{16.5} + e_{32.5})\rangle$$
$$= \langle(e_1e_{16.5} + e_2e_{16.5} + \cdots + e_{16}e_{16.5}) + (e_1e_{32.5} + e_2e_{32.5} + \cdots + e_{16}e_{32.5})\rangle.$$

Each of the 32 terms is the response of a two-element interferometer; thus the term $\langle e_i, e_j\rangle$ is the response of a two-element interferometer of spacing $(j - i)\,a/\lambda$, where λ is the wavelength and a is the unit of x expressed in the same units as for λ. Let $\psi = \pi as/\lambda$ where $s = \sin\theta$ as on p. 277. Then

$$\langle e_i e_j\rangle \propto \cos[(j - i)\Psi],$$

as may be deduced directly from the angular spectrum of two narrow antennas spaced $(j - i)a$ apart. The total response is thus

$$(\cos\psi + \cos 3\psi + \ldots + \cos 31\psi) + (\cos 33\psi + \ldots + \cos 63\psi).$$

We recognize this expression as a Fourier series; the usual formula is

$$2\sum_{n=1,3,5\ldots}^{N-1} \cos n\Psi = \frac{\sin N\Psi}{\sin\Psi} \qquad (N \text{ even}). \tag{1}$$

For N = 64 the width to half response is given by $\Delta\Psi = 0.059$ or $\Delta s = 0.019\lambda/a$.

A single antenna of length $47a/\lambda$ has a power response $\operatorname{sinc}^2(48as/\lambda)$; since $\operatorname{sinc}^2 0.44 = 0.5$, the width of half response is $\Delta s = 0.018\lambda/a$. Thus it is interesting to note that the interferometer has a beamwidth similar to that of a conventional array about half as long again. This fact can be associated with the level transfer function of the interferometer contrasted with redundant emphasis on low spatial frequencies by conventional antennas.

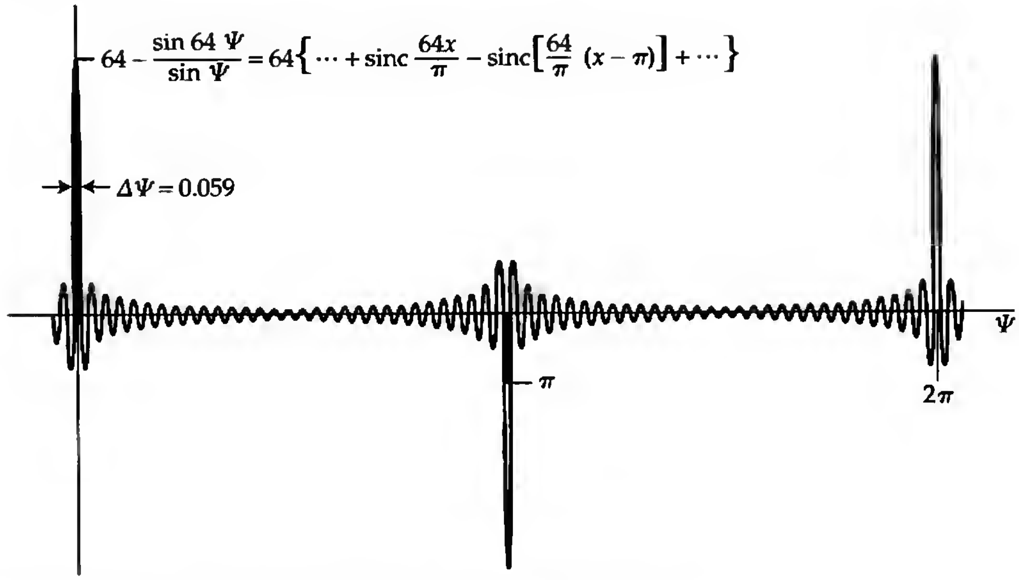


Fig. 21.15.21. The field pattern of the 18-element array.

The antenna array described was demonstrated with a spacing of 83.68 wavelengths at a wavelength of 9.107 cm, and achieved a beamwidth of 52 arcseconds (*Astrophys. J.*, vol. 138, pp. 305–309, 1963). This was the first time that an antenna reached the angular resolution of the human eye. The breakthrough was based on a novel technique of equalizing transmission line lengths to geodetic accuracy, about 1 in 10^5 (*IRE Trans. Ant. Prop.*, vol. AP-9, pp. 22–30; 75–81, 1961).

The approach to Fourier series of p. 208 (and Fig. 10.12) may be applied to the present periodic function as follows.

$$2\sum_{n=1,3,5\ldots}^{N-1} \cos nx \supset \sum_{n=1,3\ldots}^{N-1} [\delta(s + n/2\pi) + \delta(s - n/2\pi)]$$
$$= \pi\text{III}\left[\pi\left(s - \frac{1}{2}\pi\right)\right] \Pi(\pi s/N).$$

Noting that, by the inverse shift theorem,

$$e^{ix}\text{III}(x/\pi) \supset \pi\text{III}\left[\pi\left(s - \frac{1}{2}\pi\right)\right]$$

and that $e^{ix} = (-1)^m$ at $x = m\pi$, which is where $\text{III}(x/\pi)$ has its impulses, we see that

$$(2)[e^{ix}\,\text{III}(x/\pi)] * [(N/\pi)\,\text{sinc}\,(Nx/\pi)] = (-1)^m N \sum_{m=-\infty}^{\infty} \text{sinc}[(N/\pi)(x - m\pi)]$$

transforms into (2), by the convolution theorem. The form (2), a set of narrow sinc functions of alternating sign, is an alternative to the compact form (1). The latter may be derived by substituting $2\cos\, n\,x = e^{ix} + e^{-ix}$ in (1) and summing the N terms of the resulting geometric series.

24. (*a*) Let $E(x/\lambda, y/\lambda)$ represent the aperture distribution, whose FT is $P(l,m)$, and let (R,S) be a coordinate system in the same plane rotated counter clockwise through an an-

gle α with respect to $(x/\lambda, y/\lambda)$. Let $g_\alpha(R) = \int_{-\infty}^{\infty} E(x/\lambda, y/\lambda)dS$. The one-dimensional FT of $g_\alpha(R)$ gives a cross-section through $P(l,m)$ in the α-direction. Let us compare the principal cross-sections obtained as FTs of $g_0(R)$ and $g_{\pi/2}(R)$:

$$P(l,0) = \sqrt{3}\,\mathrm{sinc}^2\, l \tag{1}$$

$$P(0,m) = \frac{\sqrt{3}}{2}[\mathrm{sinc}^2\,\sigma + (i\pi\sigma)^{-1}(1 - \cos\pi\sigma\ \mathrm{sinc}\,\sigma)] \tag{2}$$

where $\sigma = (\sqrt{3}/2)m$. Let $P(l_0,0) = |P(0,m_0)| = 0.707$. Then we find that $l_0 = 0.502$ and $m_0 = 0.492$. Thus the half-power contour is highly circular (circular within one percent). The contours within the half-power contour will be even more closely circular in spite of the triangular shape of the aperture. Clearly the central curvature of $P(l,m)$ does not depend strongly on direction.

(*b*) According to the second moment theorem (p. 339), the central curvature of the two-dimensional transform is proportional to the second moment of the original function. So we can study the nature of $P(l,m)$ near its origin by examining the second moment of the triangular aperture. If we translate the problem into mechanics it becomes the question of the moment of inertia I_α of a triangular lamina about an axis lying in the plane of the triangle and passing through its centroid in a general direction α. The moment of inertia I_α of a flat plate depends on position angle α of the axis in such a way that if I_α is plotted against α in polar coordinates the result is a central ellipse. [If α is measured from a principal axis of inertia the relation is $I_\alpha = I_{xx}\cos^2\alpha + I_{yy}\sin^2\alpha$.] The only ellipse with three axes of symmetry, which is what an equilateral triangle demands, is a circle. Hence the moment of inertia, or second moment, of the triangle is independent of direction. Consequently the central curvature of $P(l,m)$ is strictly independent of direction. This is one way of understanding why the contours surrounding the origin are approximately circular.

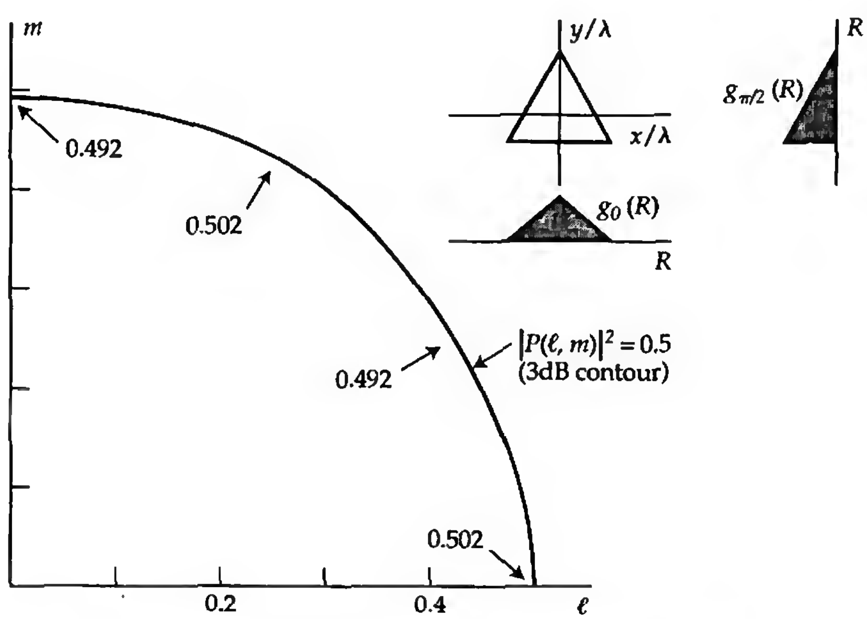


Fig. 21.15.24 **The 3dB contour of the radiation pattern of a triangular aperture is almost circular.**

Well away from the origin the pattern $|P(l,m)|^2$ must be periodic in angle with a period of 60°, i.e., it possesses six-fold symmetry. The reason for this is that the power radiation pattern, in any plane through the beam axis, is an even function. We can see from Fig. 1 that the minor departures from circularity are in the form of six bumps spaced 60°. Numbers show radial distances.

(*c*) From (1) we see that $P(l,0)$ has nulls at $l = \pm 1, \pm 2$, etc. But from (2) we see that there are no nulls on the m-axis, because at the points where the real part is zero the imaginary part is non-zero and at the point ($m = 0$) where the imaginary part is zero the real part is non-zero. Hence $|P(0,m)|^2$, which is the sum of the squares of the real and imaginary parts, has no zeros at all. We can therefore say that there is no null contour surrounding the main beam. (This does not say that there is no null contour. However at the moment we cannot say whether the nulls on the l-axis are isolated points, or lie on closed loops or lie on loci that run out to infinity.)

26. With a spacing L_λ between elements the aperture distribution will be $[\delta(x_\lambda + \frac{1}{2}L_\lambda) + \delta(x_\lambda - \frac{1}{2}L_\lambda] * E_0(x_\lambda)$. So the field radiation pattern will be proportional to $\sin \pi L_\lambda s$ multiplied by the broad angular spectrum of a single element. The innermost maxima of the power pattern will be at $s = \pm\frac{1}{2}L_\lambda$. For s to be $0.5'' = 2.4 \times 10^{-6}$ radians, means $L_\lambda = 2.1 \times 10^5$. At a wavelength $\lambda = 20$ μm the baseline length L of the interferometer would be approximately 4 m. This scheme was proposed (*Nature*, vol. 274, pp. 780–781, 1978) and used to observe the bright star Betelgeuse's immediate surroundings, which are too faint to be seen by normal astrophotography (*Nature*, vol. 395, pp. 251–253, 1998).

27. (*a*) The overall width is 10 wavelengths. (*b*) The Fourier transform of $E(\xi)$ is $(20/\pi^2 s^2)(\text{sinc } 10s - \cos 10\pi s)$. (*c*) Tabulating $P(s)$ shows a minimum of -0.0865 at $s = 0.01835$, with a corresponding dB level of $20 \log 0.0865 = 21$ dB.

■

CHAPTER 16: APPLICATIONS IN STATISTICS

3. If you purchased 10,000 ordinary 1 percent 100-ohm resistors, you could reasonably expect that their mean resistance would be between 99 and 101 ohms. You could hardly expect the manufacturer, in testing his 1 percent product, to control meteorological factors (temperature, humidity, cleanliness of contacts, supervision of technicians, absolute calibration of meters, etc.) so that his resistance measurements would be in agreement with the National Institute of Science and Technology to much better than 1 percent. It would be unreasonable if the mean resistance of your 10,000 1 percent resistors was 100.1 ohms and the total resistance of the series combination 1,001,000 ohms. If you repeated the purchase many times, buying from the same original batch, the spread could indeed possibly be as little as one part in 10^4; but about a mean that could disagree with an absolute standard by much more than one in 10^4.

8. A probability distribution $P(x)$ being positive is expressible as $|f(x)|^2$. Let $f(x)$ have a Fourier transform $F(t)$. Then if $\phi(t)$ is the Fourier transform of $P(x)$,

$$P(x) \supset \phi(t);$$

it follows from the autocorrelation theorem (p. 122) that

$$\phi(t) = F(t) \star F(t).$$

The following summarizes the relationships.

$$f(x) \supset F(t)$$
$$|f(x)|^2 \text{ or } P(x) \supset F \star F \text{ or } \phi(t).$$

A particular example is

$$\operatorname{sech} \pi x \supset \operatorname{sech} \pi t$$
$$\operatorname{sech}^2 \pi x \supset 2t \operatorname{cosech} \pi t \qquad \text{(p. 266).}$$

Thus $2t$ cosech πt is the autocorrelation function of sech πt. It is also the autocorrelation function of sech $[\pi(t-1)]$, of $\mathscr{F}\{\operatorname{sech} x \operatorname{sgn} x\}$ and in general of $\mathscr{F}\{\operatorname{sech} x\, e^{i\theta(x)}\}$, where $\theta(x)$ is any arbitrary phase, because of the fact that $f(x)$ is not uniquely defined.

10. Under convolution variances add, but second moments about the origin do not in general add. However, when probability distributions are convolved, then the second moment about the mean is the same as the variance (because the integral of the probability distribution is unity). Therefore we shall consider the third moment about the mean since there is no reason to think that third moments in general will add.

Let the probability distributions $f(x)$ and $g(x)$ have their means at $x = 0$. Then $F(0) = G(0) = 1$ and $F'(0) = G'(0) = 0$, where F and G are the Fourier transforms of f and g. The third moment of $f(s)$ is

$$\langle x^3 \rangle_f = \int_{-\infty}^{\infty} x^3 f(x)\, dx = \frac{1}{(-i2\pi)^3} F'''(0)$$

and

$$\begin{aligned}\langle x^3 \rangle_{f*g} &= \int_{-\infty}^{\infty} x^3 (f * g)\, dx = \frac{1}{(-i2\pi)^3} (FG)'''|_0 \\ &= \frac{1}{(-i2\pi)^3} [F'''G + 3F''G' + 3F'G'' + FG''']_0 \\ &= \frac{1}{(-i2\pi)^3} [F'''(0) + G'''(0)] = \langle x^3 \rangle_f + \langle x^3 \rangle_g.\end{aligned}$$

Further problem: Do the fourth central moments add? (No).

We can show that

$$\langle x^4 \rangle_{f*g} = \langle x^4 \rangle_f + 6\langle x^2 \rangle_f \langle x^2 \rangle_g + \langle x^4 \rangle_g.$$

The fourth *cumulant* $\langle x^4 \rangle - 3\langle x^2 \rangle^2$ is additive; in fact this is the defining property of cumulants.

13. Write the Poisson distribution in delta function notation as if n were a continuous variable. It is

$$\left[\delta(x) + x\delta(x-1) + \frac{x^2}{2!}\delta(x-2) + \dots\right] e^{-x}.$$

Take the Fourier transform to get the characteristic function,

$$\phi(t) = \left[1 + xe^{it} + \frac{x^2}{2!}e^{2it} + \dots\right]e^{-x}$$
$$= \exp(xe^{it})e^{-x}$$
$$= \exp(xe^{it} - x).$$

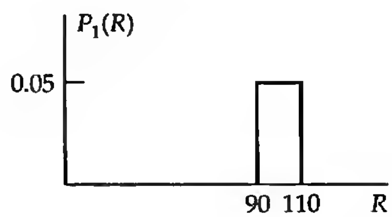
16. Let $P_1(R)dR$ be the number of resistances in $R \pm \frac{1}{2}dR$ and let $P_2(G)dG$ be the number of conductances in $G \pm \frac{1}{2}dG$. When $R = 1/G$, these numbers must be equal. In this case, $|dR| = G^{-2}|dG|$. Hence from

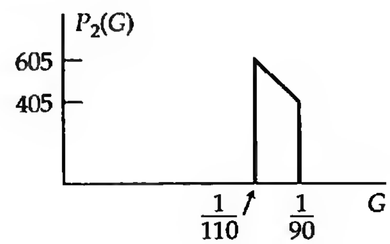
$$P_1(R)dR = P_2(G)dG$$

we deduce

$$P_2(G) = P_1(R)\left|\frac{dR}{dG}\right|$$
$$= G^{-2}P_1(R)$$
$$= 0.05\, G^{-2}\Pi\left(\frac{1}{20}G - 5\right).$$

As the graph shows $P_2(G)$ is by no means flat-topped.





CHAPTER 17: RANDOM WAVEFORMS AND NOISE

1. Let $x(t)$ have a Gaussian amplitude distribution and zero mean and a power spectrum $W(f)$. The peak-to-peak value of a long segment $x(t)\Pi(t/T) = g(t)$ is

$$\max[g(t)] - \min[g(t)]$$

and the standard deviation σ is

$$\frac{1}{T}\int_{-T/2}^{T/2}[g(t)]^2\,dt.$$

As $T \to \infty$, the first of these expressions increases without limit while the second does not. Therefore, one thing to investigate is how a factor of 5 could be found to relate two such unlike expressions.

One approach is to calculate the peak-to-peak value as a function of the segment duration T. The probability that $g(t)$ exceeds $2.5\,\sigma$ is 1 part in 161. Thus, starting from a trough where $g(t) < -2.5\sigma$, we might expect the order of 200 "cycles" to elapse before a peak occurred with $g(t) > 2.5\sigma$. By "cycle" is meant a time period $(\Delta f)^{-1}$ where

Δf, the rate of occurrence of effectively independent values, is calculable (p. 468) from the autocorrelation width of the power spectrum.

By reference to a table of probabilities for deviations from the mean of a normal distribution and rounding off heavily we find the following.

(Peak-to-peak)/σ	4	5	6	7
Number of "cycles"	50	200	1000	6000

It appears that on segments several hundred "cycles" long an accuracy better than ± 20 percent will be achieved by use of the 5 σ rule. Records for inspection by eye (and application of rules of thumb) are not likely to be much longer. Records longer than 50 cycles would yield accuracy better than 20 percent. Short records based on fewer than 50 independent values would have less accuracy, but in such cases estimates of σ are not defined to much better than 20 percent. So it seems that the 5σ rule makes sense.

Further problems suggested by this discussion are:

(*a*) Find the precise value of the factor (peak-to-peak)/σ as a function of the number of "cycles" N.

(*b*) Find a fast procedure for estimating N which does not require determination of the whole power spectrum of the sample but works instead from countable things like peaks and zero crossings.

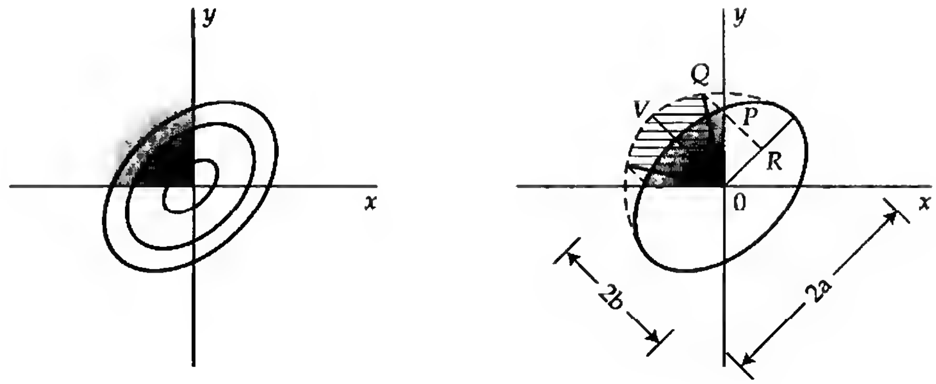
(*c*) Is the peak-to-peak value as stable a statistic as say the difference between the second highest maximum and the second lowest minimum?

2. Consider a signal waveform $H \exp(-\pi t^2)$ whose width at half peak is 0.94, and whose slope at half peak is $1.47H$. Let σ_n be the root-mean-square noise. The procedure for measuring the width is first to find the half-peak points on the given record and then to measure the time interval between them. Because of the slope, a vertical error σ_n will become a horizontal, or timing error $\sigma_n/1.47H$. The error in width, obtained by subtracting two times, each with its own independent error, will make a width error $2^{\frac{1}{2}}\sigma_n/1.47H$. Equating this to 5 percent of 0.94 gives $H/\sigma_n = 20$. This seems equitable. A more refined measuring procedure is to reduce σ_n by first convolving the record with another Gaussian waveform and then correcting for the increased width. In the theory of this procedure it is necessary to increase the factor $2^{\frac{1}{2}}$ because the half-peak timing errors come to be correlated.

4. (*a*) The expected number of upcrosses per second ν is equal to the probability that a value x_t will be negative and that x_{t+1} will be positive. Let $p(x,y)dx\,dy$ be the probability that x_t will be found at $x + \frac{1}{2}dx$ and x_{t+1} will be found at $y \pm \frac{1}{2}dy$. Put $\sigma^2 = \langle x_t^2 \rangle = \langle x_{t+1}^2 \rangle$. The correlation γ_1 between successive values is given by $\gamma_1 = \langle x_t x_{t+1} \rangle / \sigma^2$. From texts on statistics we know that $p(x,y)$ is given by the two-dimensional normal distribution

$$p(x,y) = \frac{1}{2\pi\sigma^2(1-\gamma_1^2)^{1/2}} \exp\left[-(x^2 - 2xy + y^2)/2\sigma^2(1-\gamma_1^2)\right]$$

a function which is represented by a few contours in the left figure.



So ν is equal to the shaded fraction of the volume of $p(x,y)$ in the left-hand figure or

$$\nu = \int_0^\infty dy \int_{-\infty}^0 p(x,y)dx.$$

Since the contours of $p(x,y)$ are similar and similarly situated concentric ellipses, this integral may be simplified by noting that it is equal to the shaded fraction of the area of the ellipse in the right-hand figure. The equation of the ellipse is

$$\frac{x^2}{2} - \gamma_1 xy + \frac{y^2}{2} = \text{const.}$$

Its axes lie at $\pm 45°$; substituting $y = \pm x$ gives the major and minor semiaxes:

$$a^2 = \frac{1}{1-\gamma_1}, \qquad b^2 = \frac{1}{1+\gamma_1}, \qquad a^2/b^2 = (1+\gamma_1)/(1-\gamma_1).$$

Expand the ellipse NW-SE by a factor a/b, converting it into a circle; this leaves area ratios unchanged. So ν is equal to the hatched fraction of the circle:

$$\begin{aligned} \nu &= 2V\hat{O}Q/2\pi \\ \cos 2\pi\nu &= \cos 2V\hat{O}Q \\ &= \cos^2 V\hat{O}Q - \sin^2 V\hat{O}Q \\ &= \frac{a^2-b^2}{a^2+b^2} \text{ (since } \cot V\hat{O}Q = RQ/OR = RQ/RP = a/b) \\ &= \gamma_1. \\ \therefore\ \nu &= (2\pi)^{-1} \arccos \gamma_1. \end{aligned}$$

When there is no correlation between successive terms ($\gamma_1 = 0$), the expected number of upcrosses per second is 0.25; i.e., the average period between upcrosses is 4 seconds.

(*b*) The central value of the second difference of $\{\gamma_t\}$ is $(\gamma_1 - \gamma_0) - (\gamma_0 - \gamma_1)$ or $2(\gamma_1 - 1)$ since $\gamma_0 = 1$ by normalization. Hence knowledge of γ_1 fixes the central second difference of $\{\gamma_t\}$ and vice versa. The second difference approximates to the curvature $F''(0)$ of the smooth curve when 1 is approximately equal to unity, provided $\{I_t\}$ is of such a character as to define a smooth curve. Now

$$\gamma_1 = \cos 2\pi\nu \doteqdot 1 - 2\pi^2\nu^2 + \dots .$$

Hence when $\gamma_1 \doteqdot 1$,

$$(2\pi\nu)^2 = 2(1 - \gamma_1) = -F''(0).$$

Let SS^* be the Fourier transform of the smooth curve defined by unnormalized $[\{I_t\} \star \{I_t\}]$, whose central curvature is $F''(0)\int SS^*\,df$. Then

$$\int f^2\, SS^*\, df = -(2\pi)^{-2}F''(0) \int SS^*\, df.$$

$$\therefore\ \nu^2 = \overline{f^2\, SS^*} = \frac{\int_{-\infty}^{\infty} f^2 S(f)S^*(f)df}{\int_{-\infty}^{\infty} S(f)S^*(f)df}.$$

The integral in the numerator would not exist if, for example,

$$SS^* = \operatorname{sinc}^2 f.$$

This power spectrum corresponds to taking running sums over a finite time interval. One might be tempted to say "white noise passed through a running integrator has an infinite number of zero crossings per second." This paradox will not in fact arise if we stick to running sums on discrete sequences. So we might content ourselves with observing that the integral in the numerator does exist if the range of f is limited to the spectral range occupied by the input signal. (It is reasonable to restrict the range of integration in this way since the output is unaffected by the filter transfer function in bands where there is no input.) Now if we apply input noise which has a flat spectrum up to a very high frequency indeed then the output will be considerably smoothed, but zero crossings will survive the smoothing in the sense that if we double the upper frequency limit the output zeros will remain in the same proportion to the input zeros. If we claim to put in an infinite number of zeros ("real white noise") we get out an infinite number.

This is not unreasonable behavior on the part of the filter. However, with a filter whose transfer falls off more rapidly at high frequencies (or with the filter postulated if we allow a non-zero rise time) zero crossings will be lost in the smoothing to the extent that they approach a finite limit per unit of time as the upper frequency of the flat input spectrum increases without limit.

(*c*) Maxima occur at the upcrosses of the derivative. The power spectrum of the derivative is $(2\pi f)^2$ times the power spectrum of the waveform. So the number of maxima per second is

$$\frac{\int f^4 SS^*\, df}{\int f^2 SS^*\, df}.$$

The integral in the numerator shows that maxima will be more numerous than upcrosses, and may indeed be infinite while the number of upcrosses is finite. This may be handled along lines indicated in the previous part. But even if we choose a filter whose power spectrum falls off as f^{-20}, there will still be infinite numbers of some high-order derivative, which is just as paradoxical as an infinite number of maxima. Presumably we can always get rid of these difficulties (*a*) by not ask-

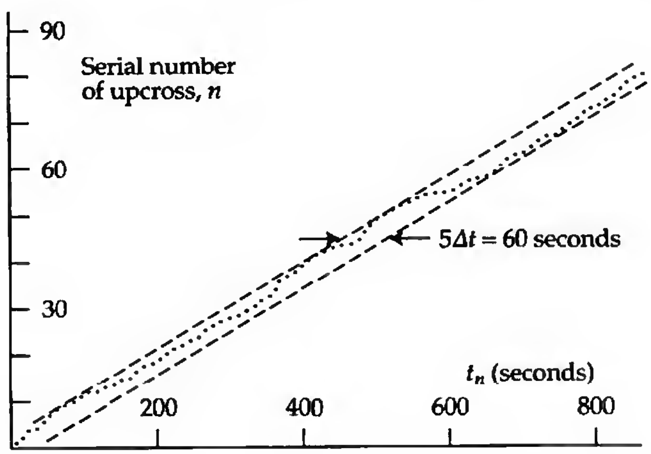
ing more of the output than we specify at the input, and (*b*) by not specifying impulse responses outside the realm of the physically possible.

5. (*a*) The sequence {1 5 10 10 5 1}, in which the successive values are taken to be one second apart, has a variance σ^2 of 1.25 seconds. Convolution with this sequence acts like a Gaussian transfer function $32(2\pi)^{1/2}\sigma\exp(-2\pi^2\sigma^2 f^2)$ since $\exp(-t^2/2\sigma^2) \supset (2\pi)^{1/2}\sigma\exp(-2\pi^2\sigma^2 f^2)$ (Prob. 1(i), p. 130). The power transfer function is proportional to $\exp(-4\pi^2\sigma^2 f^2) = \exp(-f^2/2f_0^2)$ where $f_0^2 = (4\pi^2\sigma^2)^{-1}$. Hence $f_0 = 0.142$ Hz.

   (*b*) The record of Fig. 21.17.5 has a total duration of 860 seconds and counting gives the total number of upcrosses as 82. The measured upcross frequency ν is therefore 0.095 upcrosses per second. The precision of this result is of some interest because measuring ν on a noise sample is a practical thing to do either by hand or by computer. First we compare against the theory of Prob. 16.4. The expected correlation γ_1 is determined from the autocorrelation of {1 5 10 10 5 1} at unit shift: $\gamma_1 = (5\times 1 + 10\times 5 + 10\times 10 + 5\times 10 + 1\times 5) \div (1 + 25 + 100 + .100 + 25 + 1) = 0.833$. Then

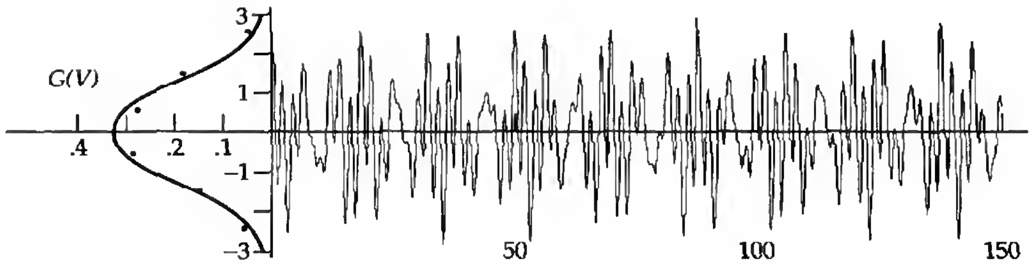
$$\nu = (2\pi)^{-1}\arccos 0.833 = 0.93\ s^{-1}.$$

   This agreement to within 2 percent is quite impressive. The precision may be investigated empirically to gain a feel for the length of record that might be required by plotting the serial number n of each upcross against its time of occurrence t_n. Using the method of Prob. 17.1 we read off Δt, the r.m.s. fluctuation of t_n from the trend value $\nu^{-1}n$ as 12 seconds. Then $\Delta\nu = (\Delta t/t_{max})\nu$ is the r.m.s. error in ν and $\Delta\nu/\nu = 12/860 = 1.4$ percent. This calculation, which is very rough, shows how the general order of precision can be arrived at empirically in the absence of prior knowledge (such as the spectrum, in this example).



8. Here we take a waveform $x(t)$ and explicitly replace it by $\mathrm{sgn}[x(t)]$. Hence the autocorrelation of the clipped signal is the same as the ψ coefficient of $x(t)$ and $x(t+\tau)$ as in Prob. 17.7.

9. The following graph shows $V(t) = \cos t + \cos et + \cos \pi t$ from $t = 0$ to $t = 150$, plotted at intervals of $t = 0.25$. The general character is certainly reminiscent of low-pass noise.



The amplitude distribution of cos kt is $P(V) = \pi^{-1}(1 - V^2)^{-1/2}\Pi(V/2)$, regardless of k, and since the three terms are uncorrelated, we expect the sum of three such variables to have an amplitude distribution $[P(V)]^{*3}$. The variance of $P(V)$, given by $V^2 = \int_{-1}^{1} V^2 P(V)dV$ is 0.5. Hence, the variance of $[P(V)]^{*3}$ is 1.5. From the central limit theorem we might expect that $[P(V)]^{*3}$ is approximated by $(2\pi)^{-1/2}(1.5)^{-1/2}\exp(-V^2/3) = G(V)$, a curve which is shown above in lieu of $[P(V)]^{*3}$, which is harder to calculate. The actual fraction of values of $V(t)$ in each amplitude range is also shown. The agreement is reasonable, but clearly a longer sample of $V(t)$ would be required before a distribution smooth enough to discriminate between $G(V)$ and some different smooth curve would result. There are 59 upcrosses. From the variance $(2\pi)^{-2}[(1 + e^2 + \pi^2)/3] = 0.154$ we expect $(0.154)^{1/2} = 0.392$ upcrosses per unit of t, or a total of 59 in $[0,150]$. Thus in certain ways the waveform simulates random noise. The peculiar spectrum could easily be revealed even in the short sample shown, but of course random noise may have any spectrum. The waveform would fail any test of randomness that brought out the coherence in narrow bands around angular frequencies of 1, e or π.

10. The real part of $\langle V_1 V_2^* \rangle$ is given by time-averaging the product of the two instantaneous voltages, and the imaginary part is obtained by introducing a quarter-cycle phase shift between the two voltages before applying them to a second multiplier and integrator. Let $F(l,m)$ be the field phasor in the direction (l,m) on a remote transverse plane. $T(l,m)$, the temperature in the direction (l,m), is proportional to $\langle FF^* \rangle$. Let $E(x/\lambda, y/\lambda)$ be the associated electric field distribution over the receiving plane, which is also transverse to the direction of the antenna beam. E and F would be a two-dimensional Fourier transform pair if the fields were steady alternating fields, and they are in this case too, where F fluctuates with time, if allowance is made for propagation time by defining E to be measured with a delay equal to the propagation time between the two planes. Thus

$$E(x/\lambda, y/\lambda) \propto \iint F(l,m)\, e^{-i2\pi(lx/\lambda + my/\lambda)} dl dm.$$

The spatial autocorrelation of E transforms into FF^*;

$$\iint E^*(x'/\lambda, y'/\lambda)E(x'/\lambda - x/\lambda, y'/\lambda - y/\lambda)d(x'/\lambda)d(y'/\lambda)$$
$$\propto \iint F(l,m)F^*(l,m)e^{-i2\pi(lx/\lambda + my/\lambda)} dl dm.$$

The expression on the left may be written $\langle E_1 E_2^* \rangle_{\text{spatial}}$, where E_1 is the field with reference to an origin which is movable with respect to E_2, whose origin is fixed. (Think of an instantaneously frozen field pattern, displace it by an amount $(x/\lambda, y/\lambda)$ with respect to itself, multiply and integrate.) Time averaging both sides,

$$\langle\langle E_1E_2^*\rangle_{\text{spatial}}\rangle \propto {}^2\mathcal{F}\{FF^*\} \propto {}^2\mathcal{F}\{T\} = \overline{T}.$$

Under circumstances where the fields at different points in the receiving plane differ in detail but have the same statistics, which will be the case for a remote constant source, spatial averaging at an instant will give the same result as time averaging at one pair of points only. Hence the time averaging on the left-hand side is sufficient. Thus

$$\langle E_1E_2^*\rangle \propto \overline{T}$$

and if both sides are normalized to be zero at their origin

$$\Gamma = \frac{\langle E_1E_2^*\rangle}{\langle E_1E_1^*\rangle}$$

$$\Gamma(x/\lambda, y/\lambda) = \frac{\langle E_1E_2^*\rangle}{\langle E_1E_1^*\rangle} = \frac{\overline{T}(l,m)}{\overline{T}(0,0)}.$$

11. The bandwidth Δf of the reception filter whose power-transfer function is $\exp(-f^2/2B^2)$ is given by $\Delta f = \frac{1}{2}W_{R\star R}$ (p. 468). The equivalent width W_R of $\exp(-f^2/2B^2)$ is $(2\pi)^{1/2}B$ (p. 59). Since variances add under convolution (p. 190) $W_{R\star R} = 2^{1/2}W_R$ and $\Delta f = 1.77B$. The integrating time of the smoothing filter is $(2\pi)^{-1/2}b^{-1}$ (Table 17.1). Hence

$$K = \frac{\text{r.m.s. fluctuation}}{\text{mean deflection}} = (\tau\Delta f)^{-1/2} = (2b/B)^{1/2}.$$

With $B = 10^4$ Hz and $b = 10^2$ Hz, $K = 0.14$. The sensitivity would have to be 14 times better to permit working to 1 percent. The integrating time $(2\pi)^{-1/2}b^{-1}$ is 4 milliseconds. Since sensitivity improves as the square root of τ it would be necessary to integrate for $14^2 \times 4$ milliseconds $= 0.8$ seconds. To establish that a change had occurred, the change would have to amount to two or three times the r.m.s. error. This would take several seconds.

Gain fluctuations of the order of 1 percent taking place in times shorter than a second could not be detected with the setup described (unless the fluctuations were systematic over considerable times), and it might not be important to know about a phenomenon simulating the generated noise. However, such gain fluctuations might be deleterious for some purposes and could be revealed by comparison of three amplifiers fed from the same noise source or by use of a single-frequency signal generator.

13. We are given that the spectrum is of the form $S(f)\Pi(f/2f_c)$, but $S(f)$ is not specified. The autocorrelation function $\langle x(t)x(t+\tau)\rangle$ is the Fourier transform of the power spectrum $SS^*\Pi(f/2f_c)$ and is therefore

$$\overline{SS^*} * (2f_c \operatorname{sinc} 2f_c t).$$

The question is whether this is zero at the sample interval $0.5\,f_c$. Now the function $(2f_c \operatorname{sinc} 2f_c\tau)$ is itself zero at $t = 0.5\,f_c$, but if we convolve this function with some other function, the zero will in general shift. Therefore, in general, critical samples are correlated.

The correlation will be zero if $\overline{SS^*} = \delta(t)$, i.e., if $SS^*(f)$ is a flat spectrum.

The average spacing of effectively independent samples must be the reciprocal of the equivalent width of the autocorrelation of $SS^*\Pi(f/2f_c)$ since the theory of pp. 466–468 shows that this is the quantity whose square root controls the increase of precision of noise power measurement.

14. Let the average field strength over an antenna aperture be E_m, but let the field depart slightly from uniform by a fraction ϵ which is a function of position in the (x,y)-plane. Then the field is $(1 + \epsilon)E_m$. The power per unit solid angle launched in the forward direction is $\iint (1 + \epsilon)E_m(1 + \epsilon^*)E_m^* dxdy$ which falls below what would be launched on axis in the absence of errors by a factor

$$\zeta = \frac{\text{aperture area}}{\iint (1 + \epsilon)(1 + \epsilon^*)dxdy}.$$

It is understood that the total radiated power is the same in both cases. If the errors ϵ are entirely due to small wavefront corrugations of amount δ, there are no amplitude variations and $\epsilon = i\delta$ and

$$\zeta = \frac{\text{aperture area}}{\iint (1 + |\epsilon|^2)dxdy} = \frac{1}{1 + \langle\delta^2\rangle}$$

where $\langle\delta^2\rangle = \iint \delta^2 dxdy/(\text{aperture area})$.

If the aperture distribution is not uniform, even in the absence of errors, as is the case for a paraboloidal reflector (because the rim is further from the focus and in addition is away from the beam maximum of the feedhorn at the focus), we introduce the concept of directivity factor $\mathfrak{D}$. For a paraboloid $\mathfrak{D}$ has a value of about 0.6 and for a rectangular feedhorn about 0.8. It represents the power reduction relative to an aperture area $\mathcal{A}$ of the same size that is excited uniformly. In this case the fractional departure from the mean is real and not particularly small and

$$\mathfrak{D} = \frac{1}{\mathcal{A}^{-1}\iint (1 + \epsilon)(1 + \epsilon^*)dxdy} = \frac{1}{1 + \text{var}\,\epsilon}.$$

If now we perturb the excitation by introducing small corrugations that bring the total errors to ϵ', we have

$$\zeta = \frac{1 + \text{var}\,\epsilon}{1 + \text{var}\,\epsilon'}.$$

Putting $\epsilon' = \epsilon + i\delta$ we get

$$\zeta = \frac{1 + \text{var}\,\epsilon}{1 + \text{var}\,\epsilon + \text{var}\,\delta}$$
$$= \frac{1}{1 + \text{var}\,\delta} = \frac{1}{1 + \delta^2}.$$

provided $\delta(x,y)$ does not correlate with $\epsilon(x,y)$, i.e., $\text{var}(\epsilon\delta) = 0$.

15. Let the N antennas be located at (x_n, y_n) where n ranges from 1 to N. Then the angular spectrum as a function of direction cosines (l, m) is (p. 418)

$$P(l,m) = N^{-1} \sum_{n=1}^{N} e^{ik(lx_n + my_n)}$$

where the factor N^{-1} has the effect of normalizing so that $P(0,0) = 1$. We could now regard x_n and y_n as random variables and treat $P(l,m)$ as a sum of random variables

and use the central limit theorem. Alternatively, impose a fine square grid of M^2 lattice points on the area G. At any lattice point (x_i,y_i) there is a probability N/M^2 of finding an antenna and a probability $1 - N/M^2$ that there is none. If there were antennas excited with amplitude a_i at each point we would have

$$P(l,m) = \left[\sum a_i\right]^{-1} \sum_{i=1}^{M^2} a_i e^{ik(lx_i+my_i)}$$

The present problem is one where a_i has a probability distribution

$$\Pr(a_i) = (1 - N/M^2)\delta(a_i) + (N/M^2)\delta(a_i - 1).$$

Using angular brackets to represent ensemble averages

$$\langle P(l,m)\rangle = \left\langle \left[\sum a_i\right]^{-1} \sum_{i=1}^{M^2} a_i e^{ik(lx_i+my_i)} \right\rangle$$

The factor $\sum a_i$ will fluctuate somewhat from arrangement to arrangement within the ensemble but will be equal to N on the average. Taking it outside the brackets and noting that $\langle a_i\rangle = N/M^2$ we get

$$\langle P(l,m)\rangle \approx N^{-1} \sum_{i=1}^{M^2} \langle a_i\rangle e^{ik(lx_i+my_i)}$$
$$= M^{-2} \sum_{i=1}^{M^2} e^{ik(lx_i+my_i)}$$

This is precisely the pattern that would result if the area G were filled with antennas. In return for the advantage of achieving with N antennas the resolution associated with M^2 antennas the price paid is a reduction of sensitivity by a factor N/M^2.

■

CHAPTER 18: HEAT CONDUCTION AND DIFFUSION

1. The differential equation governing the pouring together or confusion rather than the melting away or diffusion is obtained by changing the sign of t; thus

$$\partial^2V/\partial x^2 = -rc\,\partial V/\partial t.$$

Whether one can obtain an earlier temperature distribution by convolution is rather an interesting question because convolution tends generally to broaden and smooth a distribution whereas here narrowing and sharpening would be required. In ordinary diffusion a transfer function $\exp(-4\pi^2ts^2/rc)$ corresponding to low-pass filtering was applicable. Here it would be appropriate to use $\exp(4\pi^2ts^2/rc)$ which would have the effect of increasing high-pass filtering as we advance into the past. Any spatially Gaussian component of the temperature distribution will then revert to a point deposit of heat a finite time before. Normally one would convert a transfer function to a corresponding function of x by taking the Fourier transform. There is a serious problem of conversion of the Fourier integral in this case, but it may be tackled, with practical results, in the following way. Note that

$$\exp(4\pi^2ts^2/rc) = 1 + 4\pi^2ts^2/rc + (4\pi^2ts^2/rc)^2/2 + \cdots.$$

Transforming term by term yields

$$\delta(x) - (t/rc)\,\delta''(x) + (t/rc)^2\,\delta''''(x) - \cdots,$$

which would be the desired inverse convolving function. As a special case, if we wish to turn back the clock by only a small fraction of the time to the last deposit, only the first two terms count. Then we see that the present temperature distribution $V(x)$ is to be corrected by subtracting a small distribution that is proportional to the second derivative, since

$$[\delta(x) - (t/rc)\delta''(x)] * V(x) = V(x) - (t/rc)\,V''(x).$$

A few trials will soon reveal that we do indeed have here a convolution process that narrows and sharpens as desired.

3. Equalization is the correction of transmission system distortion due to disturbance of the relative amplitudes and phases in the Fourier components of a signal waveform. The correction is made by an equalizing filter which might, for example, compensate for reduction of low frequencies in transmission by boosting them relative to higher frequencies. Reverse diffusion could be thought of as correction for distortion of a spatial temperature pattern due to lapse of time. The analogy is rather close in that the distortion of the pattern may be compensated by a simple operation on the spatial frequencies although in this case it is the high frequencies that are to be boosted. A further resemblance is that full restoration is limited in both cases by noise. Thus, a high spatial frequency component of a temperature distribution would rapidly become small in amplitude and would then require a large restoring factor; but it would be unwise to apply the large theoretical factor if the amplitude had fallen to the level of spatial noise, or measurement error, at that spatial frequency. Reverse diffusion differs from equalization in general in not requiring phase correction and in requiring only correction factors of the form $\exp ks^2$.

4. If one-centimeter radiation fell on the moon it would be partly reflected or scattered back and partly transmitted into the interior, where it would be absorbed in the first few centimeters or so depending on the local electrical conductivity and permittivity. Conversely, radiation escaping from the moon originates in a subsurface layer, suffering attenuation appropriate to the depth of origin as it emerges. The intensity at the depth of origin is proportional to the temperature there. From p. 485, the depth of penetration d of a sinusoidal temperature variation at angular frequency ω is given by $\exp(-\alpha d) = 1/e$ where $\alpha = \operatorname{Re}\gamma = (\omega cr/2)^{-1/2}$; the lower frequencies penetrate more deeply. Consequently, below a certain depth into the moon, the temperature variation will be sinusoidal at the fundamental frequency because the harmonics will have died away. The emitted microwave radiation will be a weighted sum of contributions at frequency ω from a range of depths. The surface amplitude of the fundamental is in the neighborhood of 100 K but is modified at depth x by a factor $\exp(-\gamma x)$. This not only produces a reduction factor $\exp(-\alpha x)$ but a phase factor $\exp(-j\beta x)$. Since $\alpha = \beta$ in this case, there is a phase delay of one radian associated with each neper of reduction in amplitude. That is why the microwave full moon comes an eighth of a month late.

5. In an electric circuit $rI^2/2$ is always positive and its time integral, which represents the energy dissipated ohmically, is of necessity monotonic increasing. In heat conduction entropy is always monotonic increasing in a closed system. However, no analogy with

entropy can be found; for example, entropy may decrease at some places, but dissipated energy in a circuit cannot. Therefore, rI^2 interpreted as the product of thermal resistance per unit length and the square of heat flow has to possess some other significance. It certainly has the mathematical significance that applies to the time integral of the square of any time varying quantity, but that does not possess any specifically thermal significance. Likewise, the stored energy $CV^2/2$ in a capacitor plays an important role in circuit theory, but thermal capacity times the square of temperature is not a working concept in the theory of heat conduction. (Example: When an uncharged capacitor C is connected across an identical capacitor charged to voltage V initially containing energy $CV^2/2$ the voltage drops to $V/2$ and the energy now stored in the two capacitors is $C(V/2)^2/2 + C(V/2)^2/2$ which is only half the initial energy. This alerts us that no matter how heavy the connecting wire, half the energy is dissipated in it ohmically, *even in the limit as its resistance approaches zero*. This can be presented as a curious puzzle. However, in the thermal analogy, where a cold thin copper plate is applied face on to an identical hot plate, we understand that the temperature drops to the midvalue, and it is hard to present this behavior as puzzling even though the equations are the same as in the electrical case.)

6. The thermal capacitance of a segment dx of the bar is cdx, and if a quantity of heat dq enters the segment the temperature rise dV will be dq/cdx. Then by definition the increase in entropy of the segment will be $cdx\, dV/V$, and the increase in entropy if the temperature rises from V from an initial value T_0 to a value V_t at time t will be

$$\int_{V_0}^{V_t} cdx\, dV/V = cdx \ln(V_t/V_0).$$

The increase in entropy of the whole bar will be

$$\int_{x_{min}}^{x_{max}} c \ln(V_t/V_0)\, dx = c\int \ln V_t(x)\, dx - c\int \ln V_0(x)\, dx.$$

The final integral depends on the initial situation but is constant with time.

The integral $\int \ln V_t(x)\, dx$ is familiar in electrical transmission of images. In that interpretation x would be a two-component vector and V_t, representing intensity, would be a scalar function of the vector variable. By analogy with thermodynamics this integral is called the entropy of the image, but the resemblance is only superficial. A less misleading term would be "decibel integral." The term image entropy is also applied to $\int V \ln V\, dx$ (the reason is the resemblance to an expression $\int p \ln p\, dp$ that occurs in statistical mechanics). Needless to say, neither form of image "entropy" possesses profound properties such as increasing as time elapses.

7. Since meteors move with velocities around 40 kilometers per second, the whole of a meteor trail, which may be 10 kilometers long at a height of 80 kilometers, is formed very rapidly. We may imagine the trail of ionization created as diffusing cylindrically as time elapses. If the electron density N is to be a Gaussian function of r at each value of t, with width proportional to $t^{1/2}$, then the peak density must diminish as t^{-1} if the number of electrons is to be preserved. (In one-dimensional diffusion this factor was $t^{-1/2}$.)

The general diffusion equation (p. 475) is

$$\frac{\partial^2 N}{\partial x^2} + \frac{\partial^2 N}{\partial y^2} + \frac{\partial^2 N}{\partial z^2} = \frac{1}{K}\frac{\partial N}{\partial t}$$

and under cylindrical symmetry this becomes

$$\frac{1}{r^2}\frac{\partial^2 N}{\partial \theta^2} + \frac{\partial^2 N}{\partial z^2} = \frac{1}{K}\frac{\partial V}{\partial t}.$$

Substituting $N = (\alpha/4\pi Kt)\exp(-r^2/4Kt)$ into this differential equation we verify that we have a solution.

The refractive index is zero at frequency f in MHz when the electron density N per cubic meter is given by $N = f^2/81$. To find the maximum value of r we could differentiate $4\pi KN/\alpha = t^{-1}\exp(-r^2/4Kt)$ with respect to t and set $dr/dt = 0$, but it saves a line or two to differentiate $4\pi KN/\alpha = u\exp(-r^2u/4K)$ with respect to u and set $dr/du = 0$; thus

$$\begin{aligned} 0 &= u(d/du)(-r^2u/4K)\exp(\,) + \exp(\,) \\ &= -(u/4K)[r^2 + 2r(dr/du)u]\exp(\,) + \exp(\,). \end{aligned}$$

Setting $dr/du = 0$ we get $0 = -ur^2/4K + 1$ or

$$r^2/4Kt = 1.$$

The maximum radius thus occurs at $t = \alpha/4\pi eKN$ and the maximum radius is $(\alpha/\pi eN)^{1/2}$ in general or

$$(81\alpha/\pi ef^2)^{1/2} = 3.07\alpha^{1/2}/f$$

when $N = f^2/81$.

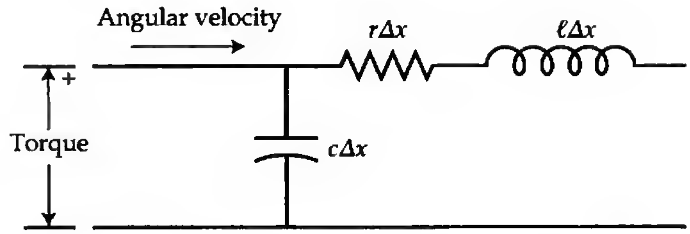
The cylinder of zero refractive index shrinks to zero radius when the central electron density $\alpha/4\pi Kt$ falls to $f^2/81$, i.e., when $t = 81\alpha/4\pi Kf^2 = 6.4\alpha/Kf^2$.

After that, the column of electrons ceases to act as a sharply bounded reflector but may continue to reflect by the mechanism of Prob. 18.8.

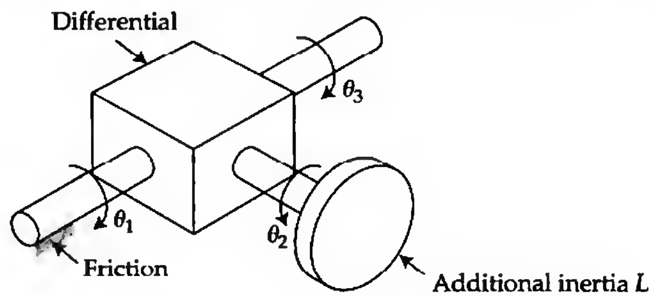
8. All the electrons lying within a range $R + x$ to $R + x + dx$ from the radio transmitter, in number $\int N(r)dy$, combine to produce an echo field strength proportional to $\exp[(i2k(R + x)]dx\,dy$ where the phase factor $\exp[(i2k(R + x)]$ allows for the total path length $2(R + x)$ and an amplitude factor depending on range has been absorbed into the constant of proportionality. Integrating over the values of x embracing the trail and taking the opportunity to absorb a phase factor $\exp(vkR)$ also we get the total echo field as proportional to $\iint N(r)\exp(i2kx)dx\,dy$. We recognize $\int N(r)dy = N_A(x)$ as an Abel transform and the expression for the total field $\int N_A(x)\exp(i2kx)dx = T(k)$ as a one-dimensional Fourier transform. With k fixed we recognize this integral as one whose maximum value (p. 174) could reach $\int N_A(x)dx = \iint N(r)dx\,dy = \alpha$ if all the electrons were confined to the axis. Consequently the power reduction factor, as the column expands, is $|\alpha^{-1}\iint N(r)\exp(i2kx)dx\,dy|^2$. The absorbed factors do not affect this conclusion.

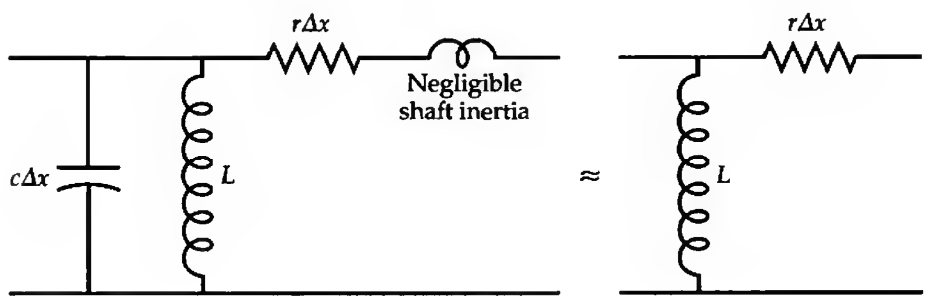
   Now, since the wavenumber k is proportional to frequency we see that the Fourier transform $T(k)$ is a transfer function describing the frequency response of the ionized column. Thus the quantity $N_A(x)$ may be regarded as the impulse response resulting from an incident impulse of electrical field. It gives the spatial waveform of the reflected field.

9. No strictly continuous system has been proposed but an analogue to the discrete circuit of Fig. 18.8 can be set up. Imagine a drive shaft turning in a viscous medium that exerts a torsional resistance r per unit length. (The unit of $r\Delta x$ is unit torque per unit angular velocity.) Let the moment of inertia per unit length be I and the torsional com-



pliance per unit length be c. Then the equivalent circuit of length Δx would be as shown. Now connect an additional inertia to each segment by means of a differential gear





inserted in the shaft as illustrated. Then the equivalent circuit would acquire an extra inductance L, where L equals the extra moment of inertia (referred to the shaft). The property of the differential is that $\theta_3 = \theta_1 - \theta_2$. Now suppose that the shaft inertia is negligible ($l\Delta x\,\omega << r\Delta x$), a condition that is met when the angular velocity is steady ($\omega = 0$) or only slowly varying.

In addition, let

$$L\omega \ll 1/\omega c \Delta x$$

which means that much more current flows through L than through $c\Delta x$. In such circumstances the combination of the two shunt elements is effectively an inductance. This new condition is similar to the first as both will be met for small enough ω. Or, of course, we can suppose that torsional compliance is negligible. A series resistance shunt-inductance circuit then results, as in Fig. 18.8. There is a dual to the mechanical scheme (imagine a massive shaft of viscous nonelastic pitch suspended by spiral clocksprings). The impossibility of converting these systems to a continuous one is apparent, and it may be possible to prove that no such continuous torsional system can exist.

However, a continuous electrical system can be conceived in the form of a coaxial transmission line whose inductive reactance per unit length is negligible compared with the series resistance of the inner conductor. The shunt inductance is introduced by filling the line with plasma and operating at frequencies (well below the plasma frequency) so that the convection current carried between the conductors by free electrons dominates the displacement current associated with the capacitance between the conductors.

13. (*a*) Sinusoidal response to sinusoidal excitation occurs in the presence of linearity plus time invariance. A few feet below the surface, only conductance can take place, so we need not take into account that heat transfer from surfaces by radiation is nonlinear, and that heat removal by convection is complicated. Because the thermal resistivity and specific heat can be treated as constants (independent of temperature) over small ranges of temperature, nonlinearity should not be important. (One may note that the range of temperature variation is only moderate and that the temperature does not go down to the freezing point of water where significant changes in soil constants might be expected.) Time invariance is a separate question since it may very well happen that the soil is wet in winter and dry in summer, but in practice distortion of the annual sinusoid is hardly noticeable at depths of a few feet.

(*b*) The characteristic thermal impedance $Z_0 = (r/j\omega c)^{1/2}$ has a phase angle of 45°. Therefore the heat flow lags the temperature variation by one-eight of a year. The temperature maximum comes one-eighth of a year after $t = 0$, so the heat flow maximum comes one-quarter of a year after $t = 0$, and is expressible by $I(t) - I_0 \cos(\omega t - \pi/2)$. Thus $\alpha = -\pi/2$.

(*c*) The diffusion equation

$$\frac{\partial^2 \theta}{\partial x^2} = rc\frac{\partial \theta}{\partial t}$$

becomes, for sinusoidal time variation,

$$\frac{d^2\theta}{dx^2} = j\omega rc\theta$$

where θ is the temperature phasor (i.e., $\theta(t) = \mathrm{Re}\,\theta e^{j\omega t}$). The solution $\theta = \theta_0 \exp[-(j\omega rc)^{1/2}x]$ shows that the amplitude dies out with depth in proportion to $\omega^{1/2}$. So the diurnal variation does not penetrate very deeply and below that the simple annual variation predominates.

14. Consider a drive shaft that transmits torque and angular motion. It is a continuous physical system and is one-dimensional in the sense that its state is adequately represented by functions of x only, where x is distance along the shaft. A suitable pair of functions might be torque $T(x)$ and angular velocity $\Omega(x)$. Let the torsional compliance per unit length be c; this means that if a length dx is clamped at one end, then $c\,dx$ is the ratio of angular displacement to applied torque. Let the torsional resistance per unit length be $r(x)$; this means that if a steady torque exists on a free section of length dx then $r\,dx$ is the ratio of that torque to the accompanying steady angular velocity. The drive shaft requires no further parameters to specify it (for example, it has no inertia). The equations are:

$$\partial T/\partial x = -r\Omega \qquad \text{and} \qquad \partial\Omega/\partial x = -c\,\partial T/\partial t.$$

Therefore, by eliminating Ω and T respectively,

$$\partial^2 T/\partial x^2 = rc\partial T/\partial t \qquad \text{and} \qquad \partial^2\Omega/\partial x^2 = rc\partial\Omega/\partial t.$$

Since inertia is zero, it might seem improbable that practical examples of inertialess driveshafts could arise in practice, but in fact we do not require absence of mass, only that inertial impedance should be negligible compared with torsional resistance. Thus, if the moment of inertia per unit length is l, we should be looking for situations where $l\omega \ll r$. This condition is more than adequately met by steady rotation and may be complied with very well where angular velocity is only slowly varying. (Remember that $\omega = 0$ for $\Omega =$ const. with time.) An example would be a drill turning in contact with the wall of its hole. The dual of this case would arise with a rotating shaft where inertia was important, but there was no external friction and negligible torsional stiffness. The second parameter would be viscous torsional conductance. (Picture a glassblower spinning a vase on the end of an overheated white-hot glass tube, so hot that viscous angular flow dominates torsional elastic deflection.)

15. Consider a vertical column of cross-section area A. Then the thermal resistance per meter to vertical flow of heat is given by $r = 1/kA$ thohms per meter, and the thermal capacitance per meter is given by $c = \rho sA$ tharads per meter (p. 476). If x is measured downward from the surface and the surface temperature phasor is $V(0)$ then the subsurface temperature $V(x)$ is given by $V(x) = V(0)\exp(-\gamma x)$, where the propagation constant $\gamma = (i\omega cr)^{1/2}$ and ω is the angular frequency to which the phasor refers (p. 485). Splitting the propagation constant into an attenuation constant α and a phase constant β we have $\gamma = \alpha + i\beta = (\omega cr/2)^{1/2} + i(\omega cr/2)^{1/2}$ and $V(x) = \exp(-\alpha x)\,V(0)\exp i\beta$. The attenuation factor in this case where $x = 1$ m, $\omega = 2\pi/86{,}400$ rad s^{-1} and $rc = \rho s/k = 2.1 \times 10^6$ s m^{-2} is given by $\exp(-\alpha x) = \exp[-(\omega cr/2)^{1/2}] = \exp(-8.5) = 2 \times 10^{-4}$. Thus the 24-hour component of the surface temperature variation is reduced by 8.5 nepers, a factor of 2×10^{-4} (and since $\alpha = \beta$, its phase is retarded by 8.5 radians or nearly 33 hours). The surface temperature does not vary sinusoidally but possesses harmonics. They will be attenuated even more, so that the temperature one meter down will be very much more sinusoidal. It would seem that one meter is much deeper than necessary for many practical purposes. For the annual temperature variation the attenuation factor is $\exp(-8.5/365^{1/2}) = \exp(-0.45) = 0.64$, so the seasonal variation is not much reduced and the lag is 0.64 radians or approximately one month.

■

CHAPTER 19: DYNAMIC SPECTRA AND WAVELETS

3. An expression meeting the description is

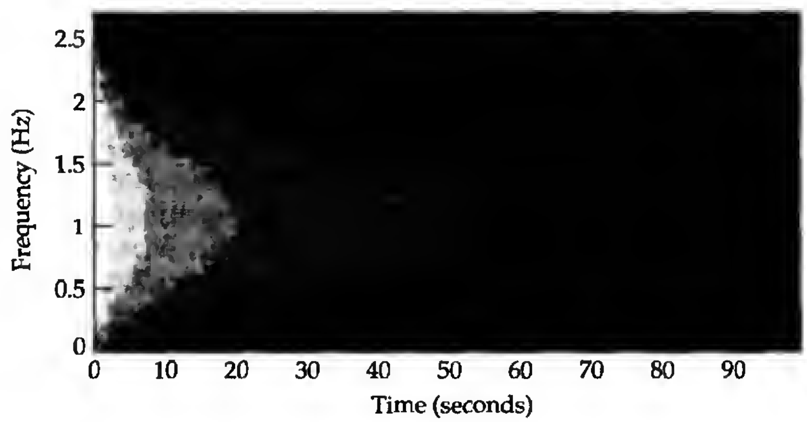
$$(t - t^2)\cos(200t + 400t^2 - 267t^3).$$

4. (*a*) The velocity of approach is $v = -gt$, taking $t = 0$ to be the moment when the train halts. The doppler shifted frequency is $550 + 550(v/c) = 550(1 - gt/c)$, where c is the speed of sound. Integrate to get the corresponding phase $550(t - gt^2/2c)$. Hence $s(t) = \cos[550(t - gt^2/2c)]$.

5. (*a*) The mean frequency is zero, as for all real signals. (*b*) The standard deviation is analogous to the radius of gyration of the line mass corresponding to

$|S(f)|^2 = 0.5 \times 10^{-5} e^{-\pi[10^5(f+60)]^2} + 0.5 \times 10^{-5} e^{-\pi[10^5(f-60)]^2}$ and, for such a narrow-band signal, is close to 60 Hz. The exact value is $[60^2 + (10^{-5}/2.5066)^2]^{1/2}$. The factor 2.5066 is the ratio of equivalent width to standard deviation.

6. Since the first zero crossing is at $t = 2.405/2\pi$, the lowest frequency to be dealt with might be judged to correspond to a period 4 times $2.405/2\pi$ or 0.65 Hz. A choice of $\Delta f = 0.08$ Hz and $\Delta t = 1$ s might be a good combination to start with. However, the lengthening of period that is apparent to the eye from a graph of $J_0(2\pi t)$, and could be supported by a table of intervals between zero crossings, is not demonstrable on a (t,f) diagram. The figure, due to David Choi, does show a reduction of amplitude as time elapses, as expected.



7. The Haar wavelet may be expressed in terms of two rectangle functions, but may also be expressed as $-\text{rect}\, x \,\text{sgn}\, x$ shifted half a unit right. The transform of $-\text{rect}\, x\, \text{sgn}\, x$ is $-\text{sinc}\, s * (-i/\pi s)$, a convolution representing the Hilbert transform of $-i\,\text{sinc}\, s$, which is known to be $i(\pi s)^{-1}(1 - \cos \pi s)$. Allow for the right shift by multiplication with $\exp(-i\pi s)$ to get $F(s) = (\pi s)^{-1}(1 - \cos \pi s)(\sin \pi s + i \cos \pi s)$.

22

Pictorial Dictionary of Fourier Transforms

Throughout the following table of Fourier transform pairs a function of x is on the left, and on the right is the minus-i Fourier transform, a function of s given by

$$\int_{-\infty}^{\infty} f(x)e^{-i2\pi sx}\,dx.$$

Small ticks show where the variables have a value of unity, imaginary quantities are shown by broken or thin lines, and impulses are shown by arrows of a length equal to the strength of the impulse.

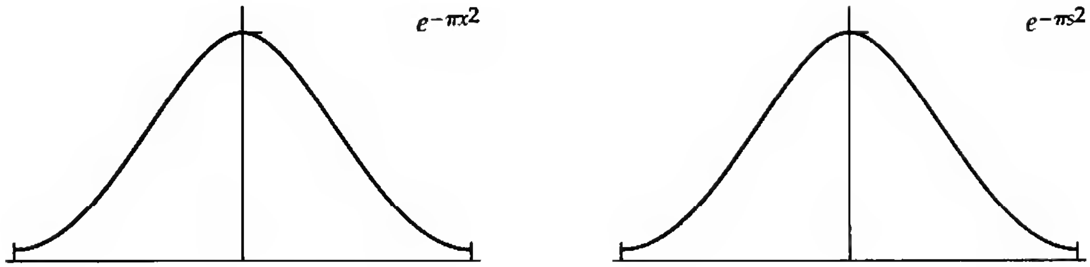
Most of the functions $f(t)$ presented have either even or odd symmetry, partly because of the inconvenience of representing a spectrum that is complex by two curves, one for the real part and one for the imaginary. Consequently, most of the Fourier transforms are also Hartley transforms. Some Hartley transforms $H(s)$ of unsymmetrical functions are given at the end; additional examples can easily be generated from a given Fourier transform $F(s) = \mathcal{R}(s) + i\mathcal{I}(s)$ using $H(s) = \mathcal{R}(s) - \mathcal{I}(s)$. In these cases, all of the information about the spectrum is contained in just the one curve.

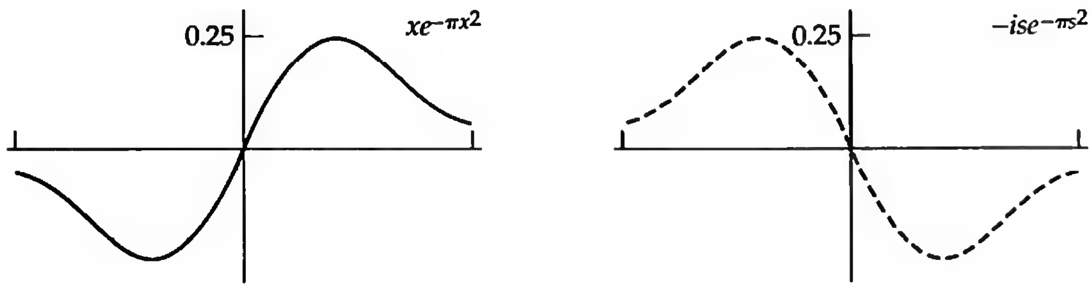
This pictorial dictionary, which contains the most frequently encountered transform pairs, is useful for finding the transform of a function, whether it is given graphically or analytically. There are, of course, many transforms that are uncommon, and it is necessary to search for these in tables.[1]

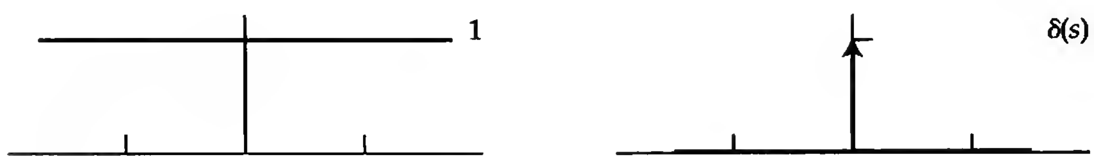
There are a number of other uses for this pictorial dictionary, such as checking numerical values in calculations and browsing through it for inspiration. One can also practice transforms by verifying that $F(0)$ equals the area under $f(x)$ and/

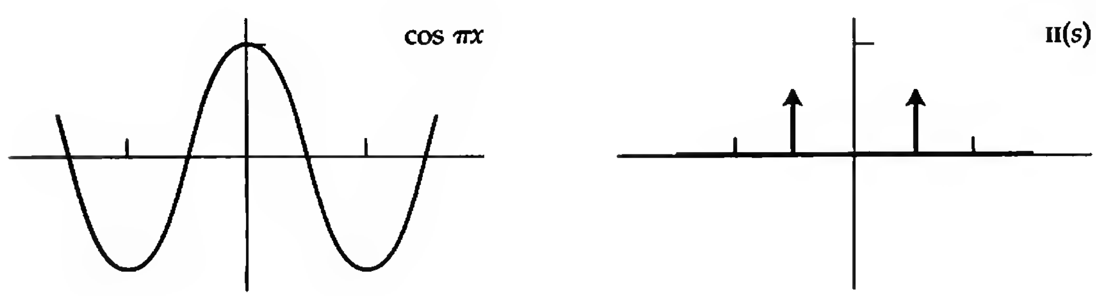
[1] G. A. Campbell and R. M. Foster, *Fourier Integrals for Practical Application*, John Wiley & Sons, New York, 1948; A. Erdélyi (ed.), *Tables of Integral Transforms*, McGraw-Hill Book Company, New York, 1954.

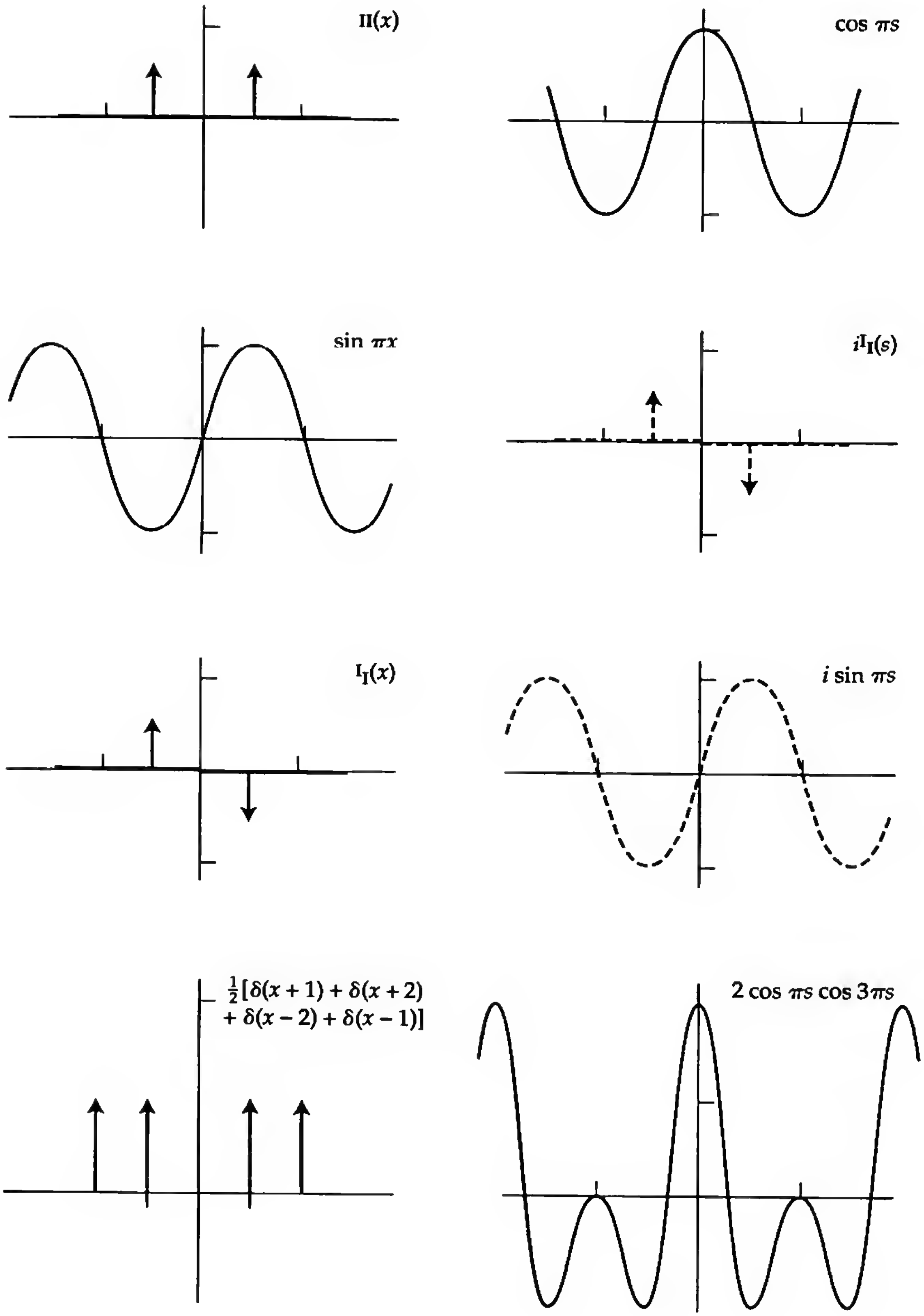
or that $f(0)$ equals the area under $F(s)$, at least approximately. If this quick test is passed, there may still be a sign error; check this by comparing the sign of the first moment of $f(x)$ with the slope of $F(s)$ at the origin; the signs should be opposite. Corresponding checks may also be applied to transform pairs generated from discrete samples; when a packaged program is first used such checks are essential to make sure that the sampling intervals Δx and Δs are compatible, to ascertain whether it is the direct or inverse transform, if either, that has a multiplier N^{-1}, to see what convention has been followed regarding 2π, and to check that it is the direct transform that uses $-i$.

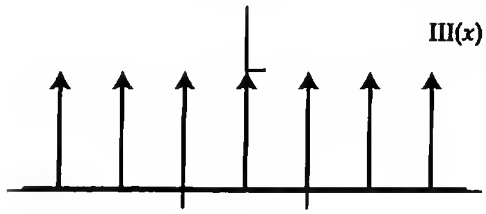

$e^{-\pi x^2}$
$e^{-\pi s^2}$

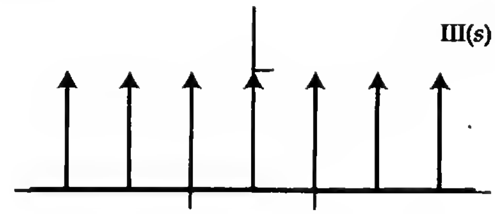

0.25
$xe^{-\pi x^2}$
0.25
$-ise^{-\pi s^2}$

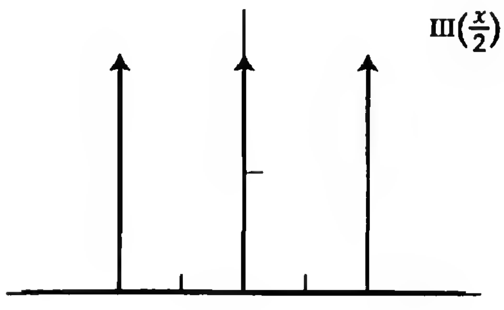

1
$\delta(s)$

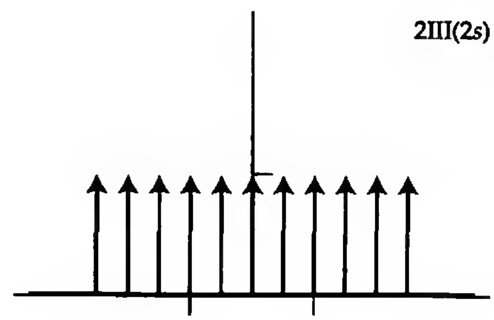

$\cos \pi x$
$\text{II}(s)$

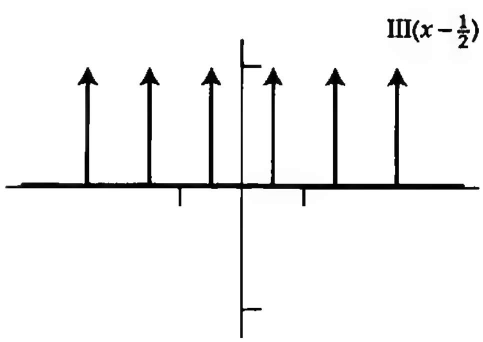

$\mathrm{II}(x)$
$\cos \pi s$
$\sin \pi x$
$i\mathrm{I_I}(s)$
$\mathrm{I_I}(x)$
$i \sin \pi s$
$\frac{1}{2}[\delta(x+1)+\delta(x+2)$
$+\delta(x-2)+\delta(x-1)]$
$2 \cos \pi s \cos 3\pi s$

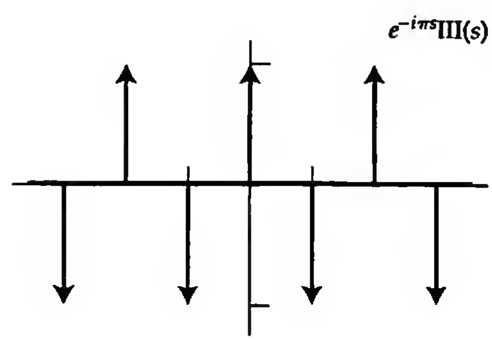

$\mathrm{III}(x)$

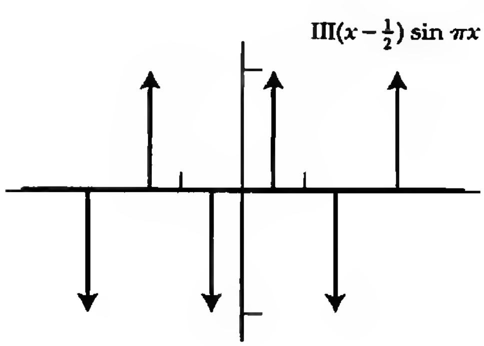

$\mathrm{III}(s)$

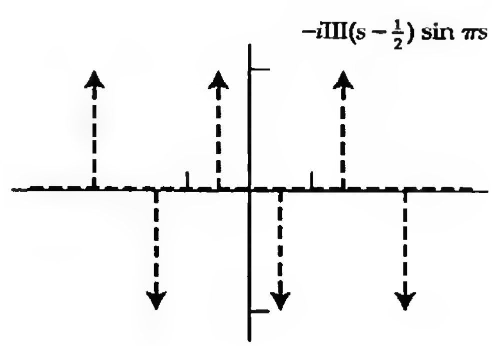

$\mathrm{III}\left(\frac{x}{2}\right)$

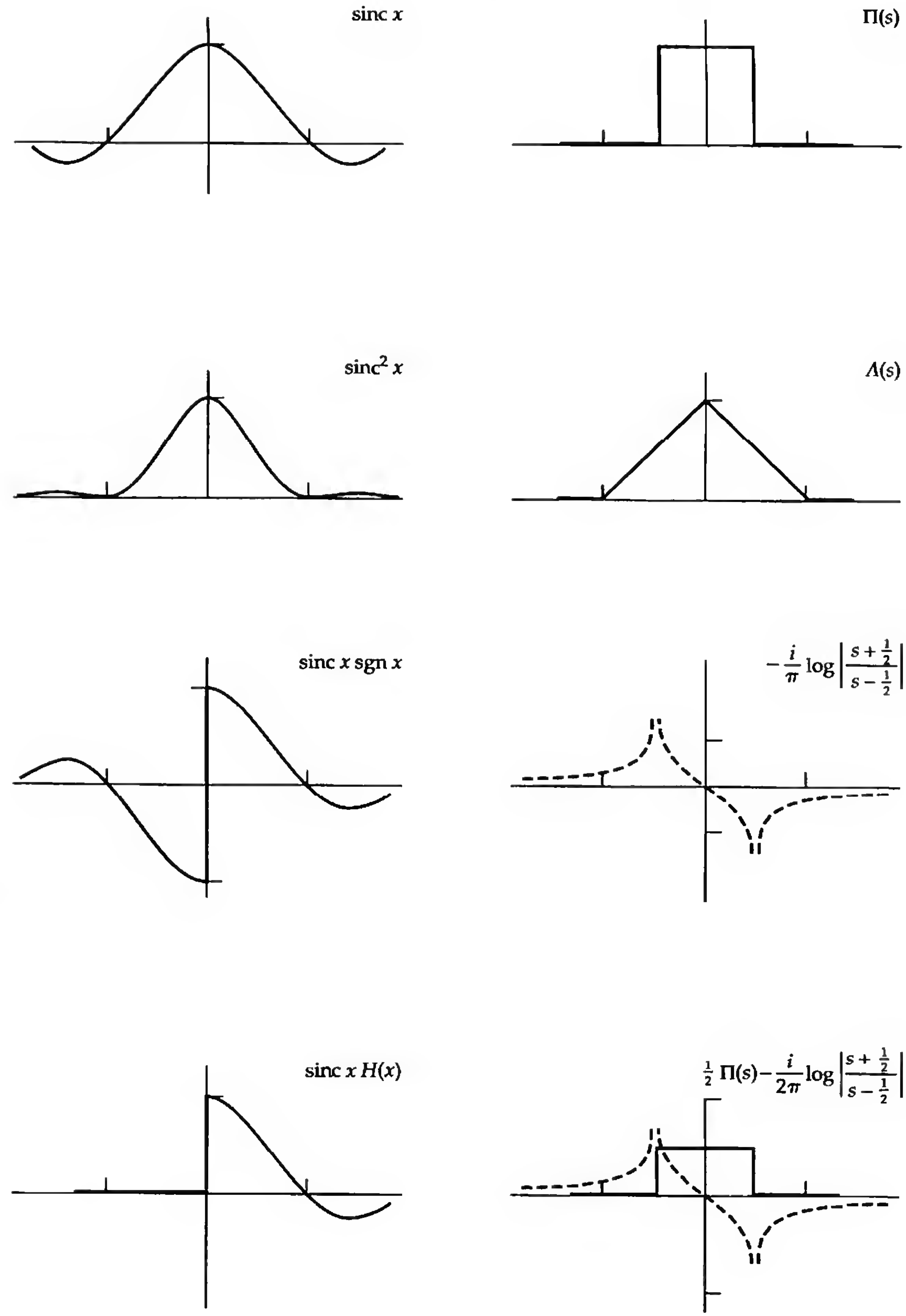

$2\mathrm{III}(2s)$

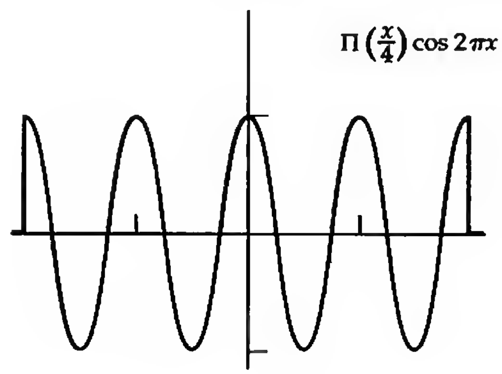

$\mathrm{III}(x-\frac{1}{2})$

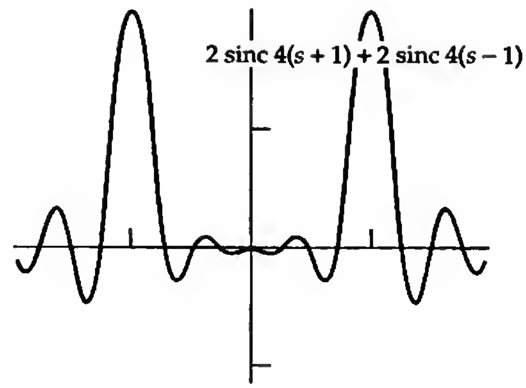

$e^{-i\pi s}\mathrm{III}(s)$

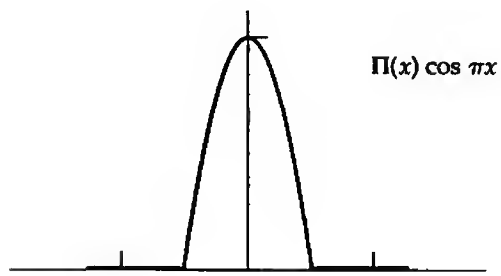

$\mathrm{III}(x-\frac{1}{2})\sin\pi x$

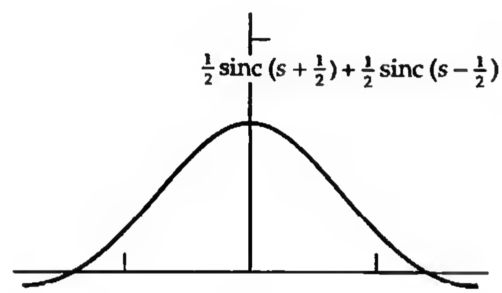

$-i\mathrm{III}(s-\frac{1}{2})\sin\pi s$

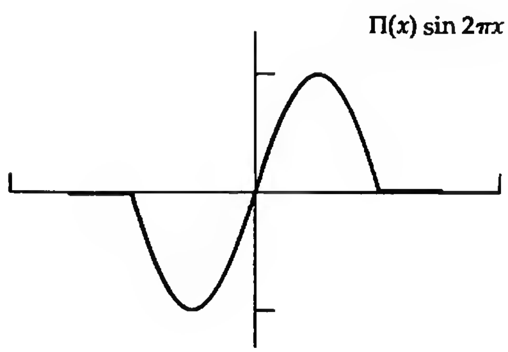

sinc x
$\Pi(s)$
$\operatorname{sinc}^2 x$
$\Lambda(s)$
sinc x sgn x
$-\frac{i}{\pi}\log\left|\frac{s+\frac{1}{2}}{s-\frac{1}{2}}\right|$
sinc $x\, H(x)$
$\frac{1}{2}\Pi(s)-\frac{i}{2\pi}\log\left|\frac{s+\frac{1}{2}}{s-\frac{1}{2}}\right|$

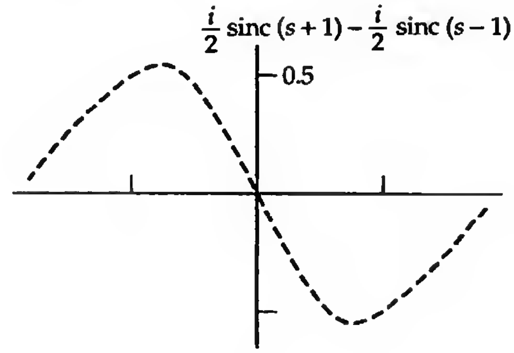

$\Pi\left(\frac{x}{4}\right)\cos 2\pi x$

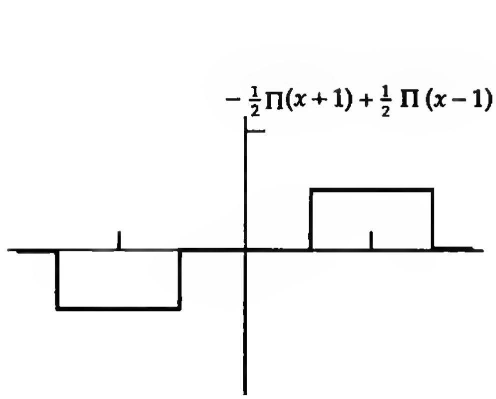

$2\operatorname{sinc} 4(s+1) + 2\operatorname{sinc} 4(s-1)$

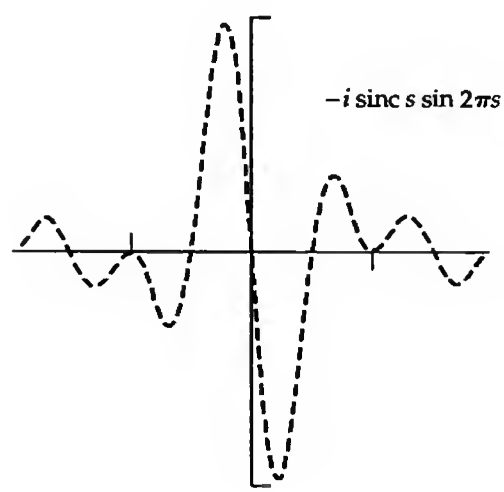

$\Pi(x)\cos \pi x$

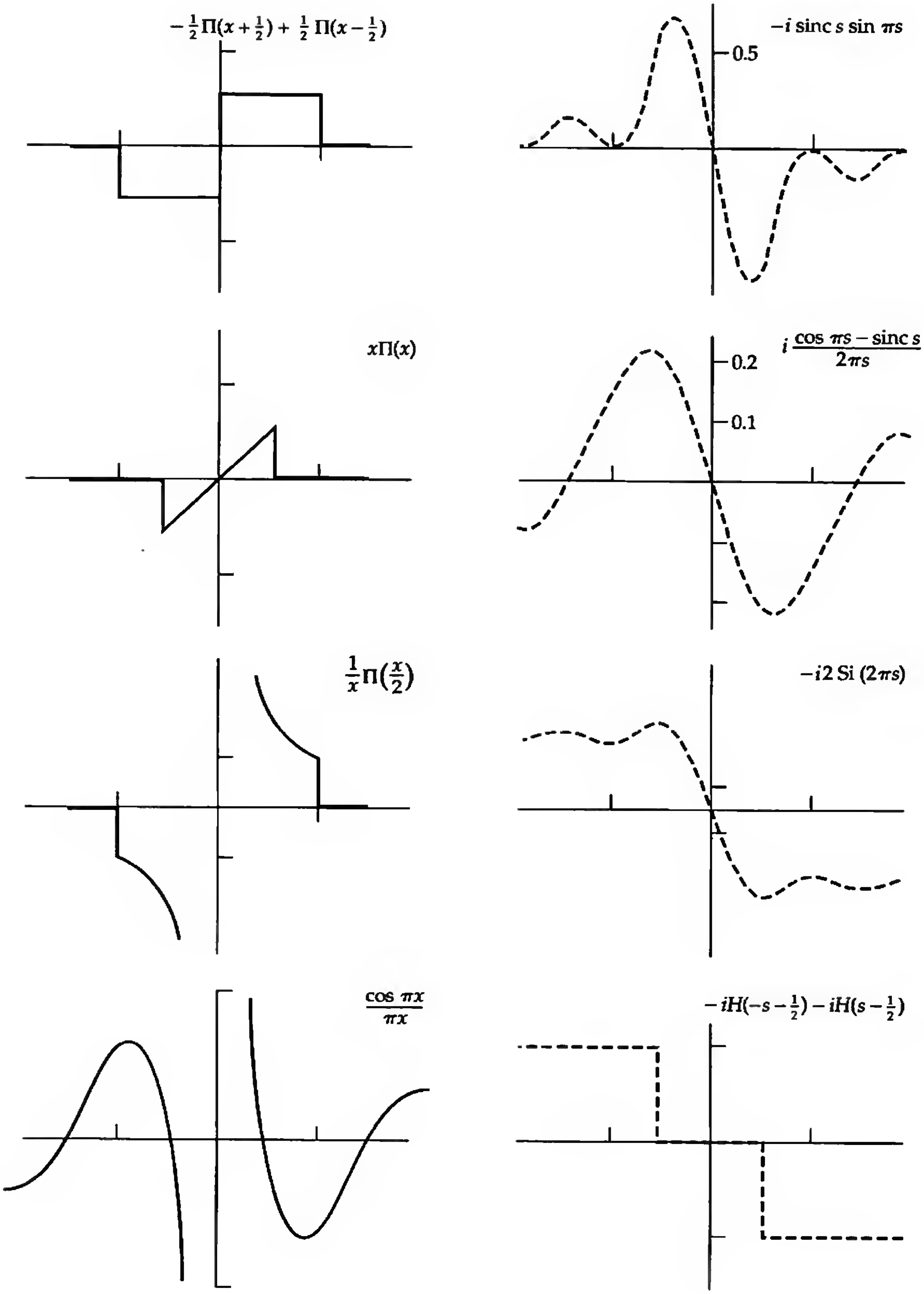

$\frac{1}{2}\operatorname{sinc}(s+\frac{1}{2}) + \frac{1}{2}\operatorname{sinc}(s-\frac{1}{2})$


$\Pi(x)\sin 2\pi x$

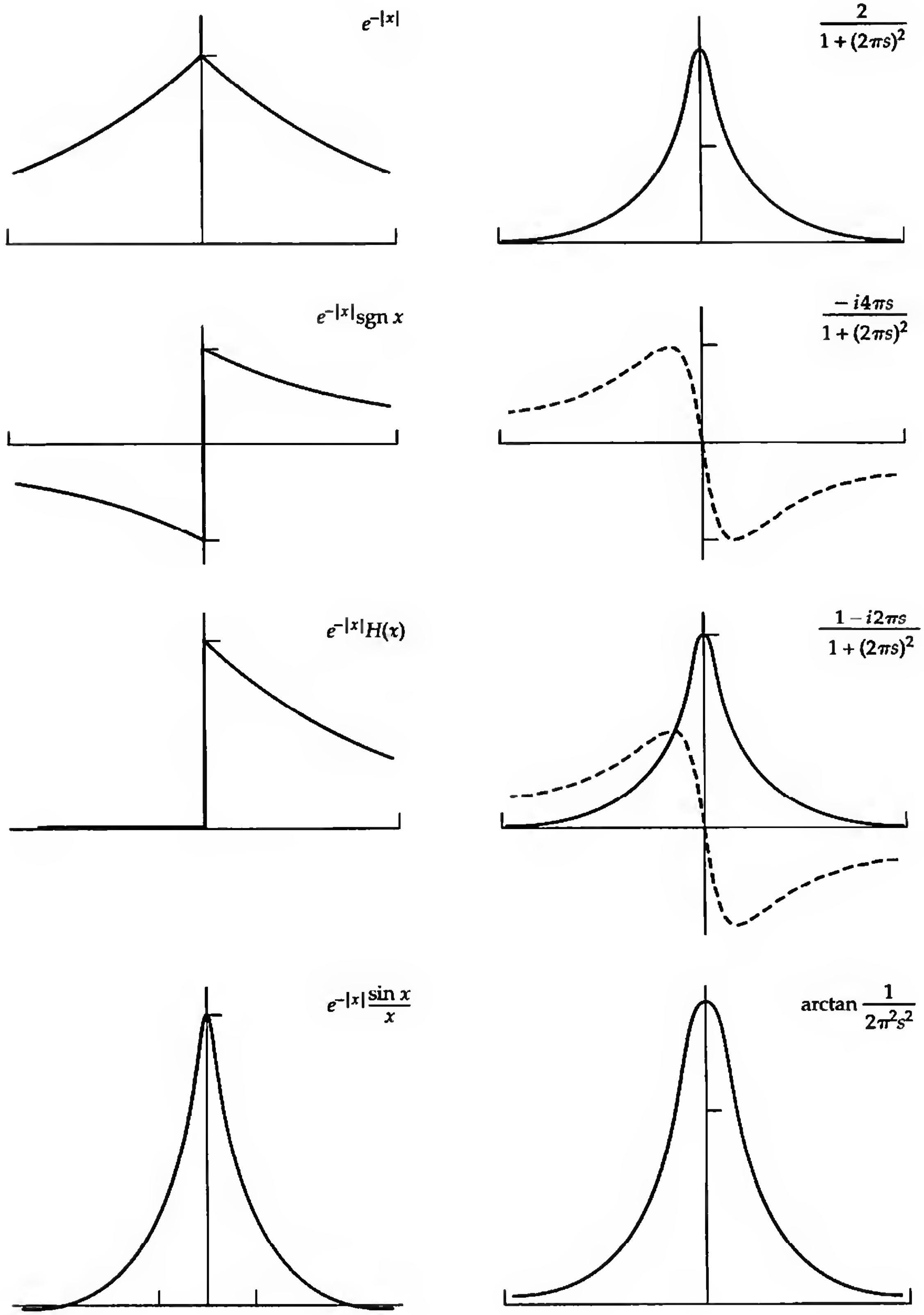

$\frac{i}{2}\operatorname{sinc}(s+1) - \frac{i}{2}\operatorname{sinc}(s-1)$
0.5

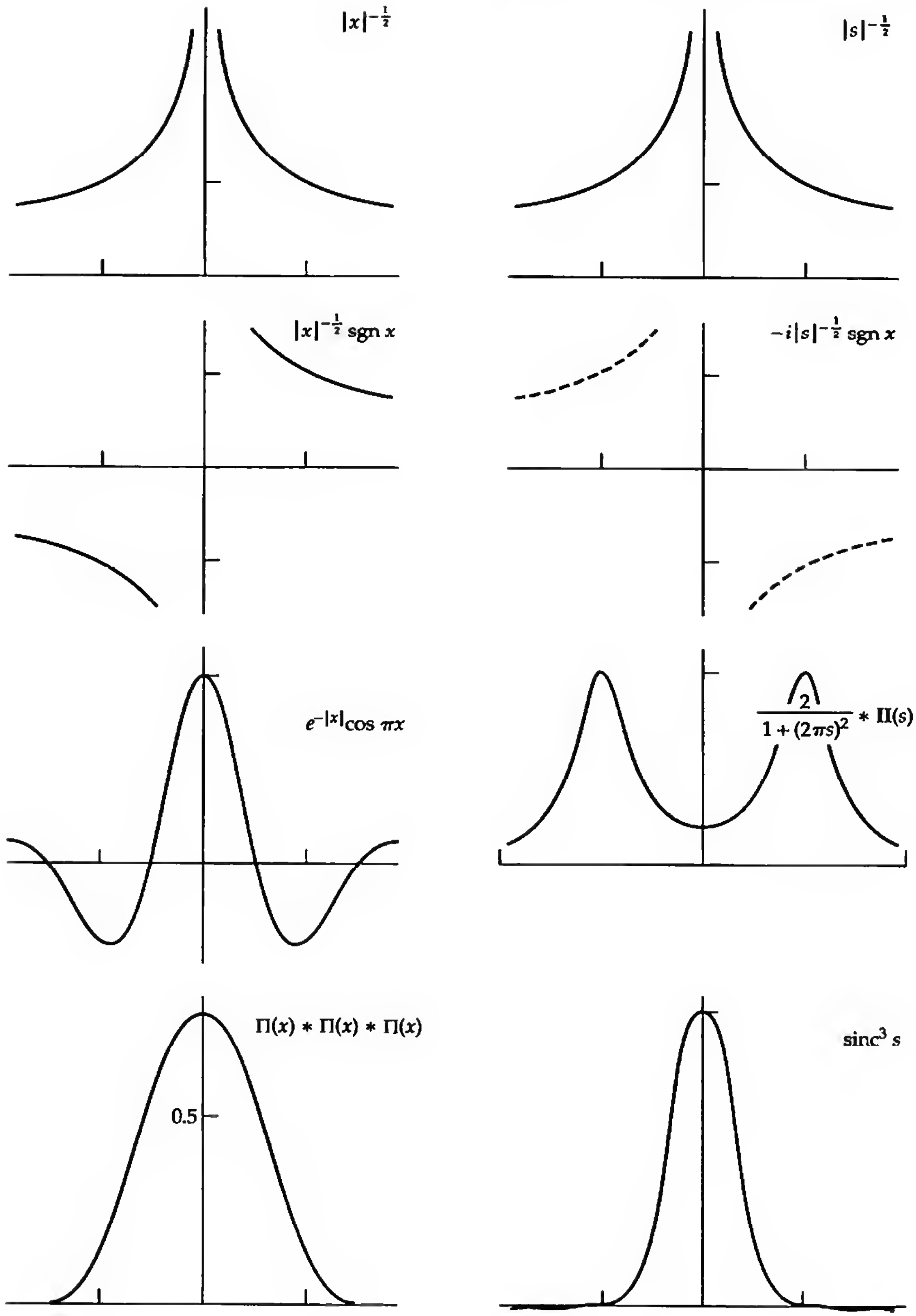

$-\frac{1}{2}\Pi(x+1) + \frac{1}{2}\Pi(x-1)$

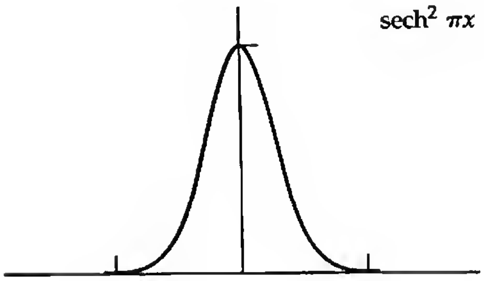

$-i\operatorname{sinc} s \sin 2\pi s$

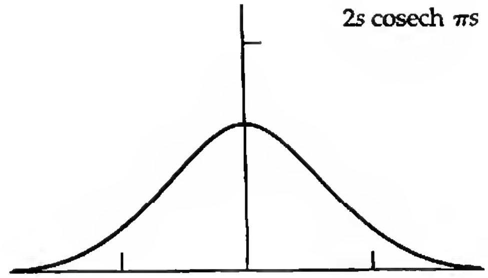


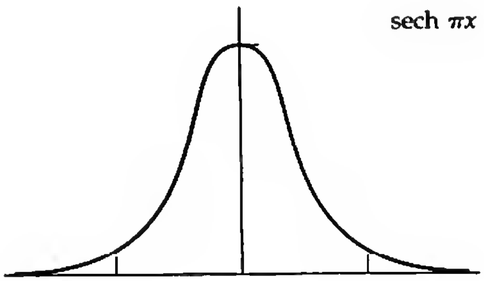
$-\frac{1}{2}\Pi(x+\frac{1}{2}) + \frac{1}{2}\Pi(x-\frac{1}{2})$
$-i\,\text{sinc}\,s\,\sin\pi s$
0.5
$x\Pi(x)$
$i\,\dfrac{\cos\pi s - \text{sinc}\,s}{2\pi s}$
0.2
0.1
$\frac{1}{x}\Pi(\frac{x}{2})$
$-i2\,\text{Si}\,(2\pi s)$
$\dfrac{\cos\pi x}{\pi x}$
$-iH(-s-\frac{1}{2}) - iH(s-\frac{1}{2})$

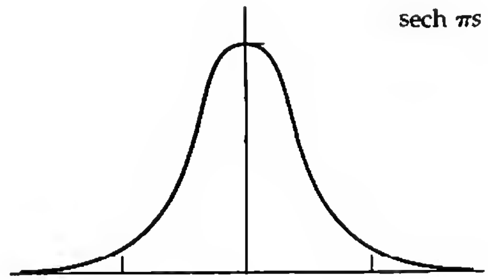

$e^{-|x|}$
$\frac{2}{1+(2\pi s)^2}$
$e^{-|x|}\operatorname{sgn} x$
$\frac{-i4\pi s}{1+(2\pi s)^2}$
$e^{-|x|}H(x)$
$\frac{1-i2\pi s}{1+(2\pi s)^2}$
$e^{-|x|}\frac{\sin x}{x}$
$\arctan \frac{1}{2\pi^2 s^2}$

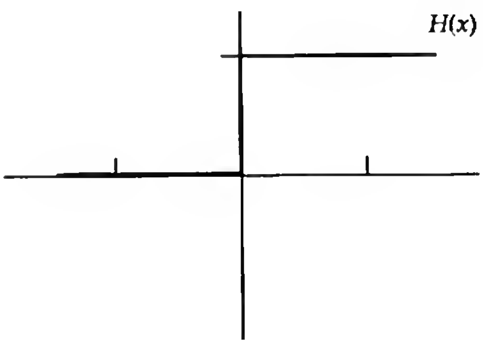

$|x|^{-\frac{1}{2}}$
$|s|^{-\frac{1}{2}}$
$|x|^{-\frac{1}{2}} \operatorname{sgn} x$
$-i|s|^{-\frac{1}{2}} \operatorname{sgn} x$
$e^{-|x|} \cos \pi x$
$\frac{2}{1+(2\pi s)^2} * \Pi(s)$
$\Pi(x) * \Pi(x) * \Pi(x)$
0.5
$\operatorname{sinc}^3 s$

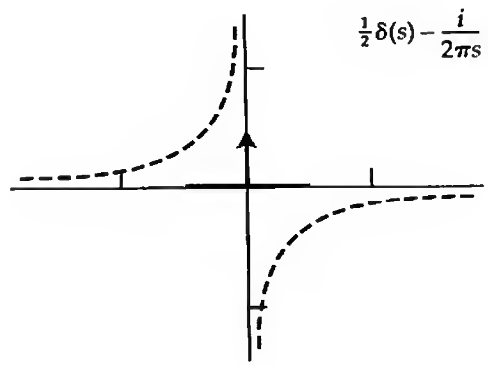

$\operatorname{sech}^2 \pi x$

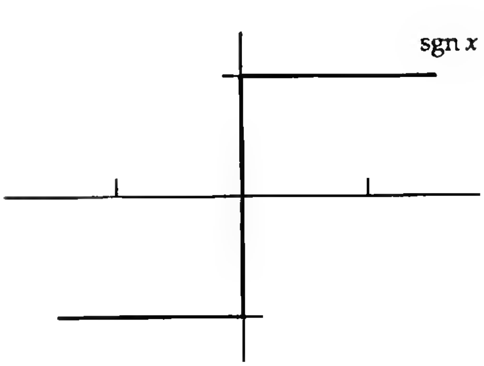

$2s \operatorname{cosech} \pi s$

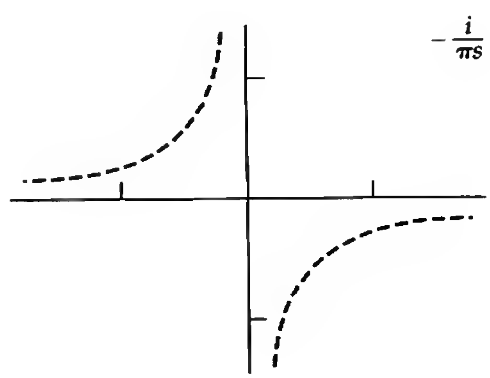

$\operatorname{sech} \pi x$

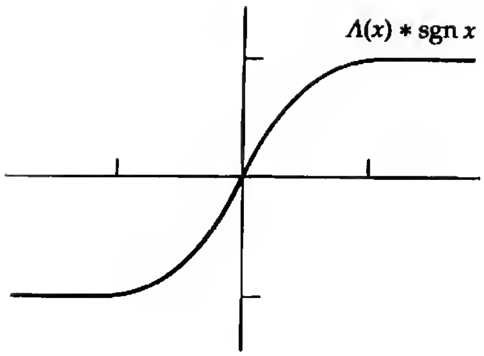

$\operatorname{sech} \pi s$

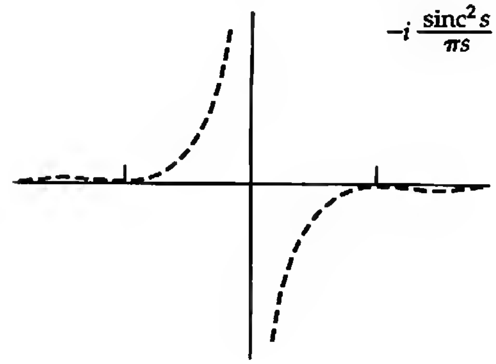

$H(x)$

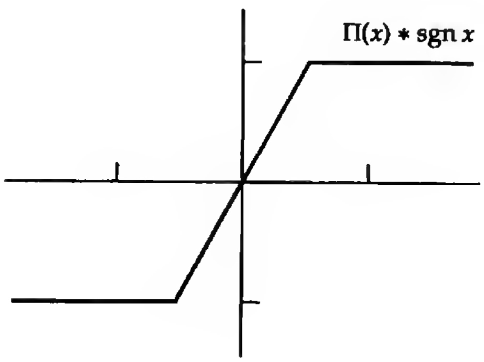

$\frac{1}{2}\delta(s) - \frac{i}{2\pi s}$

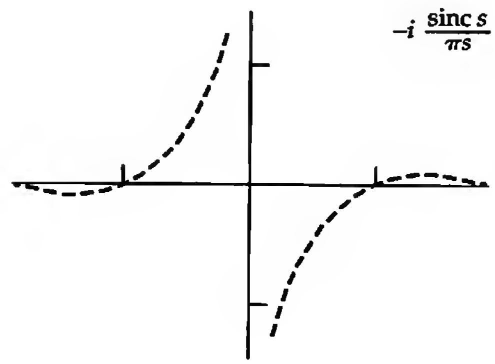

$\operatorname{sgn} x$

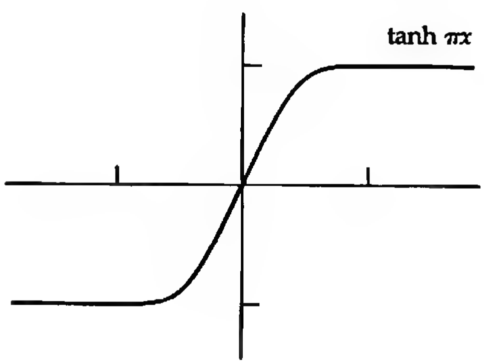

$-\frac{i}{\pi s}$

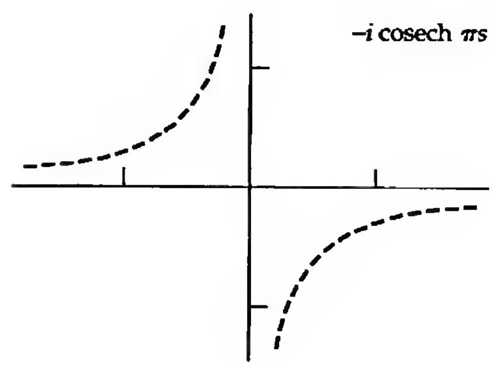

$\Lambda(x) * \operatorname{sgn} x$

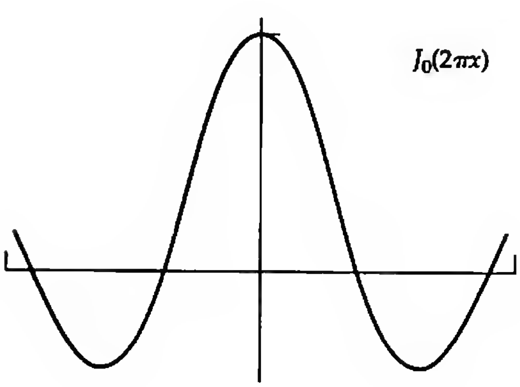

$-i\,\dfrac{\operatorname{sinc}^2 s}{\pi s}$

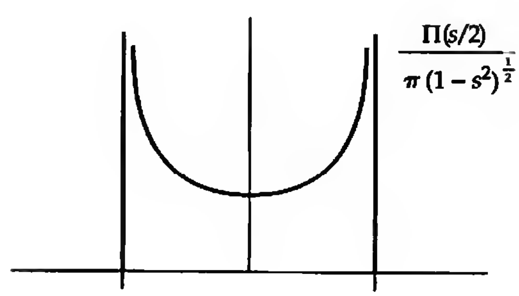

$\Pi(x) * \operatorname{sgn} x$

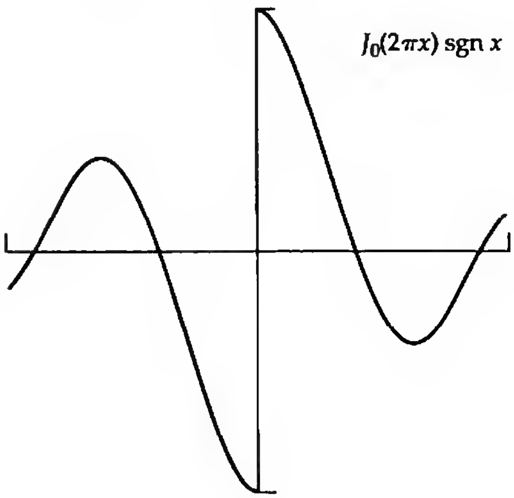

$-i\,\dfrac{\operatorname{sinc} s}{\pi s}$

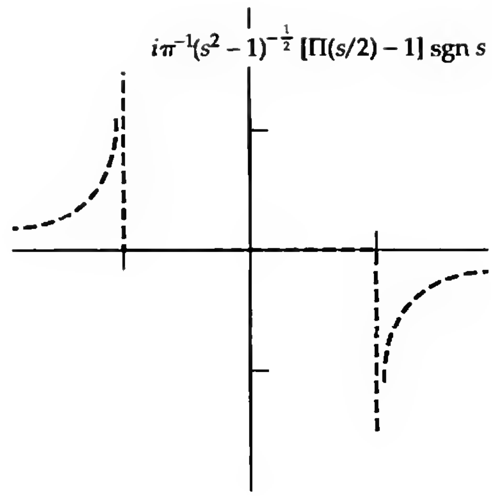

$\tanh \pi x$

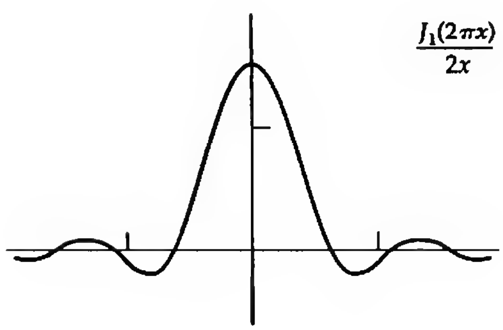

$-i \operatorname{cosech} \pi s$

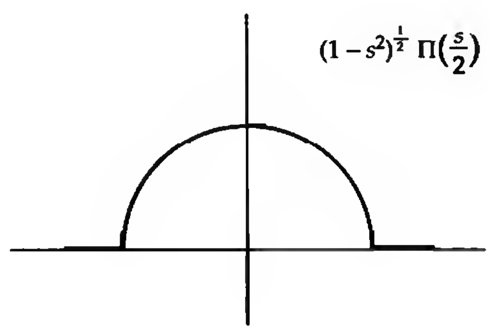

$J_0(2\pi x)$

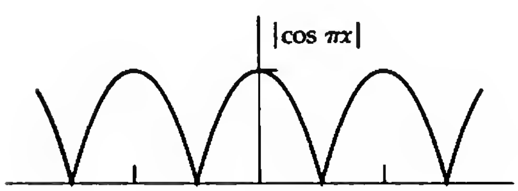

$\dfrac{\Pi(s/2)}{\pi(1-s^2)^{\frac{1}{2}}}$

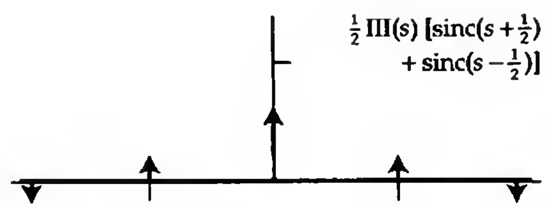

$J_0(2\pi x)\,\mathrm{sgn}\,x$

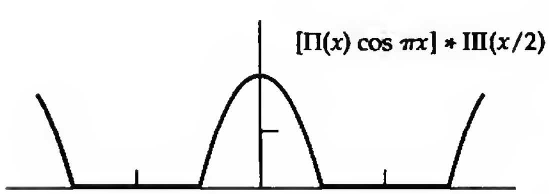

$i\pi^{-1}(s^2-1)^{-\frac{1}{2}}\,[\Pi(s/2)-1]\,\mathrm{sgn}\,s$

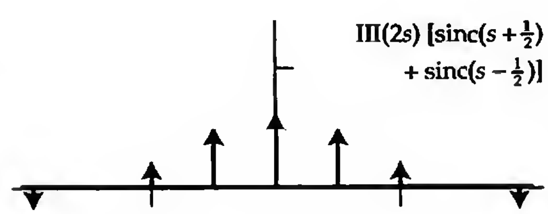

$\dfrac{J_1(2\pi x)}{2x}$

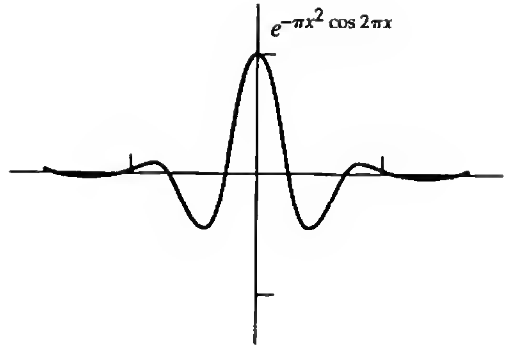

$(1-s^2)^{\frac{1}{2}}\,\Pi\left(\frac{s}{2}\right)$

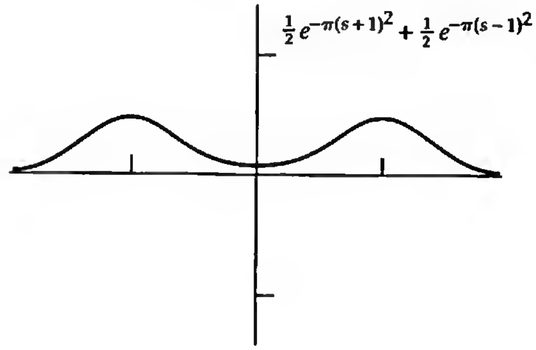

$|\cos \pi x|$

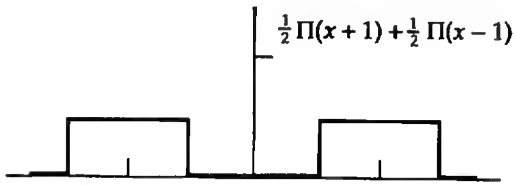

$\frac{1}{2}\,\mathrm{III}(s)\,[\mathrm{sinc}(s+\frac{1}{2})$
$+\,\mathrm{sinc}(s-\frac{1}{2})]$

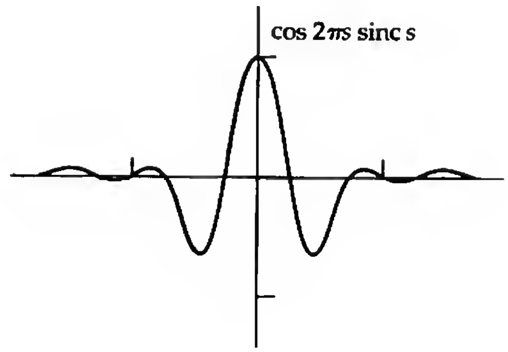

$[\Pi(x)\cos \pi x] * \mathrm{III}(x/2)$

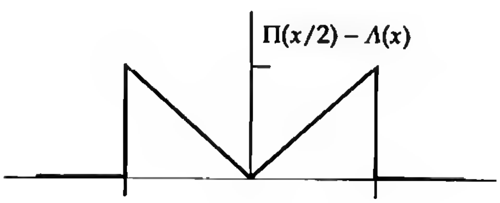

$\mathrm{III}(2s)\,[\mathrm{sinc}(s+\frac{1}{2})$
$+\,\mathrm{sinc}(s-\frac{1}{2})]$

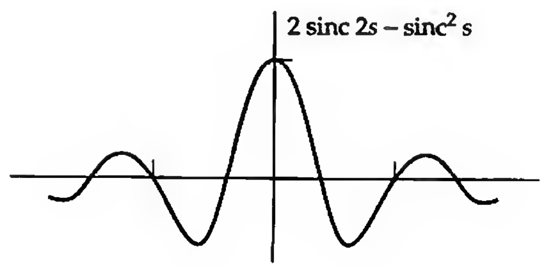

$e^{-\pi x^2} \cos 2\pi x$

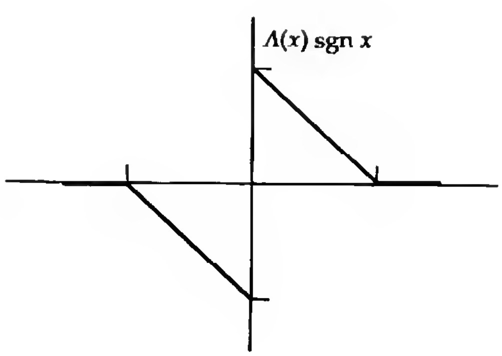

$\frac{1}{2} e^{-\pi(s+1)^2} + \frac{1}{2} e^{-\pi(s-1)^2}$

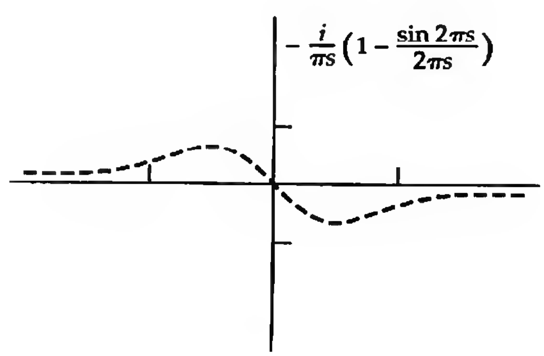

$\frac{1}{2} \Pi(x+1) + \frac{1}{2} \Pi(x-1)$

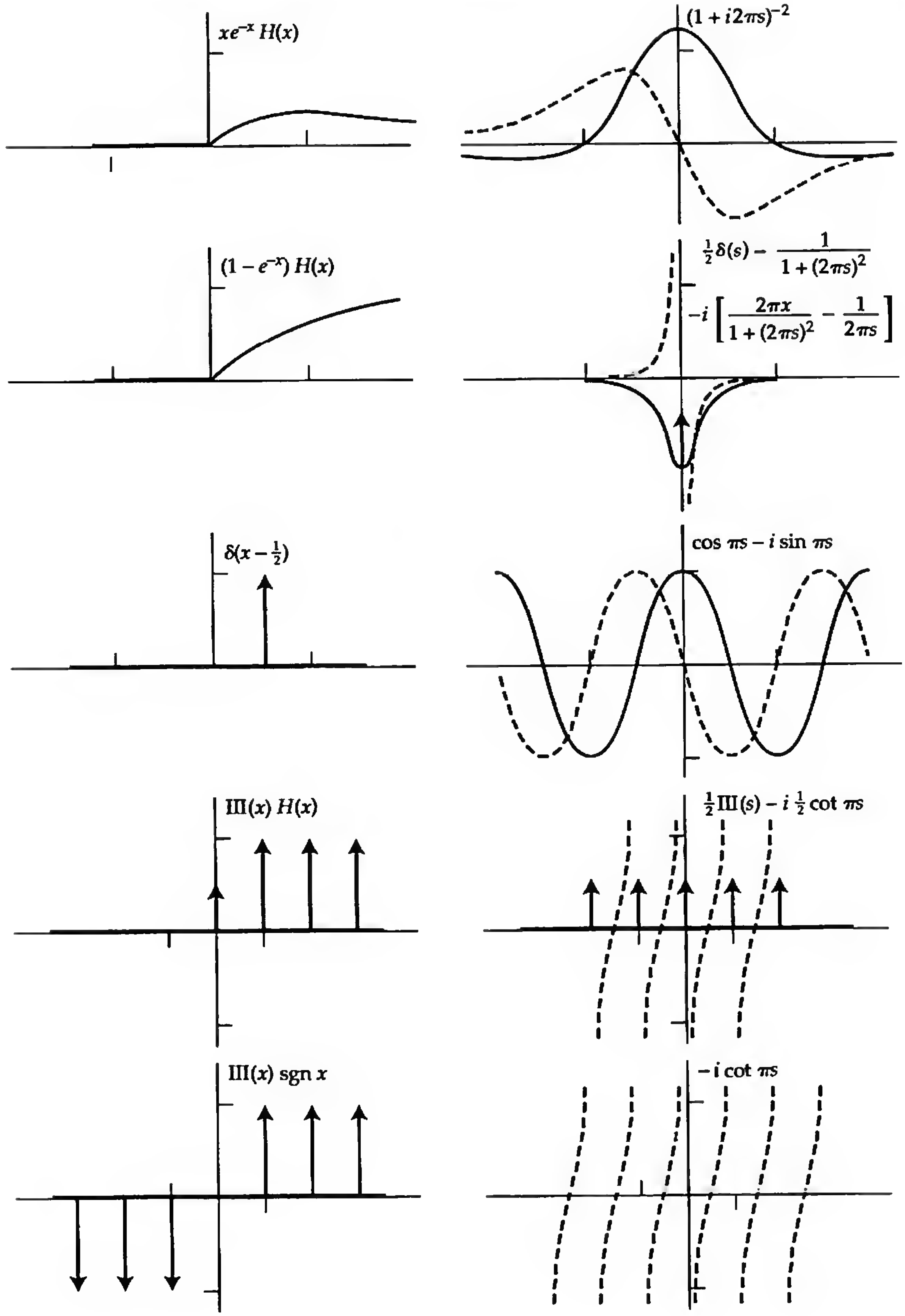

$\cos 2\pi s \operatorname{sinc} s$

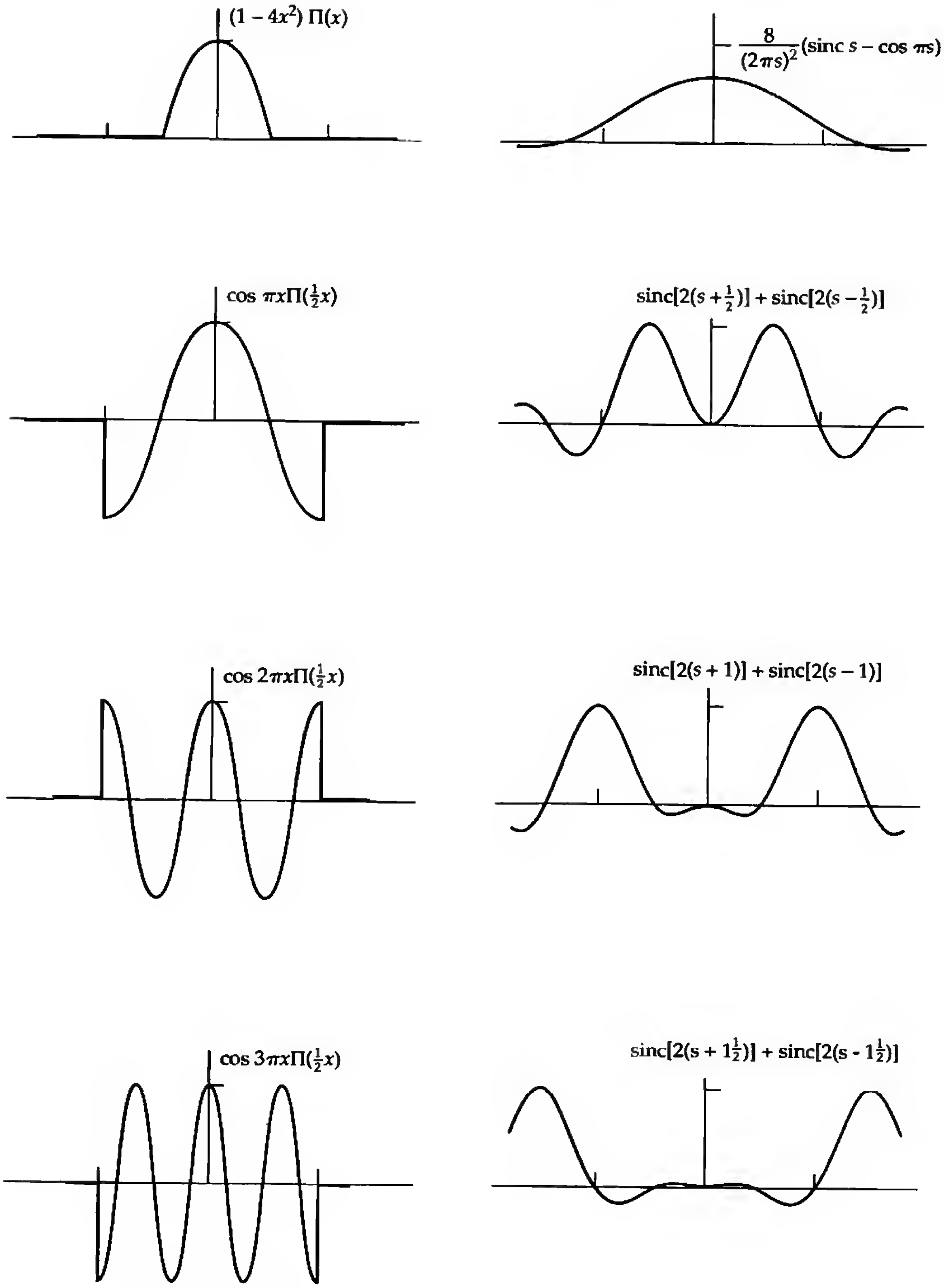

$\Pi(x/2) - \Lambda(x)$

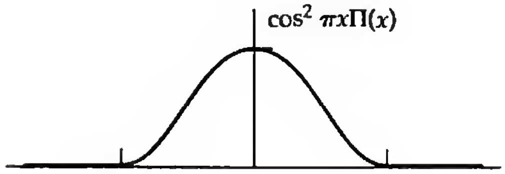

$2 \operatorname{sinc} 2s - \operatorname{sinc}^2 s$

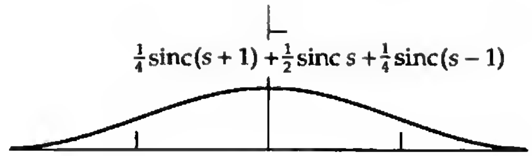

$\Lambda(x) \operatorname{sgn} x$

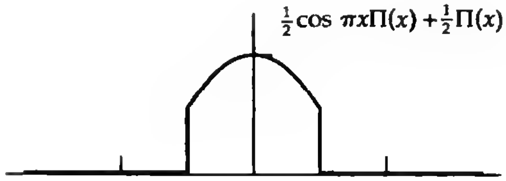

$-\frac{i}{\pi s}\left(1 - \frac{\sin 2\pi s}{2\pi s}\right)$

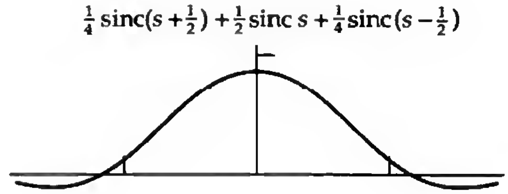

$xe^{-x} H(x)$
$(1 + i2\pi s)^{-2}$
$(1 - e^{-x}) H(x)$
$\frac{1}{2}\delta(s) - \frac{1}{1 + (2\pi s)^2}$
$-i \left[\frac{2\pi x}{1 + (2\pi s)^2} - \frac{1}{2\pi s} \right]$
$\delta(x - \frac{1}{2})$
$\cos \pi s - i \sin \pi s$
$\mathrm{III}(x)\, H(x)$
$\frac{1}{2}\mathrm{III}(s) - i\, \frac{1}{2} \cot \pi s$
$\mathrm{III}(x) \operatorname{sgn} x$
$-i \cot \pi s$

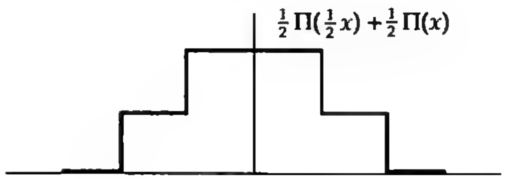

$(1-4x^2)\,\Pi(x)$
$\frac{8}{(2\pi s)^2}(\operatorname{sinc} s - \cos \pi s)$
$\cos \pi x \Pi(\frac{1}{2}x)$
$\operatorname{sinc}[2(s+\frac{1}{2})] + \operatorname{sinc}[2(s-\frac{1}{2})]$
$\cos 2\pi x \Pi(\frac{1}{2}x)$
$\operatorname{sinc}[2(s+1)] + \operatorname{sinc}[2(s-1)]$
$\cos 3\pi x \Pi(\frac{1}{2}x)$
$\operatorname{sinc}[2(s+1\frac{1}{2})] + \operatorname{sinc}[2(s-1\frac{1}{2})]$

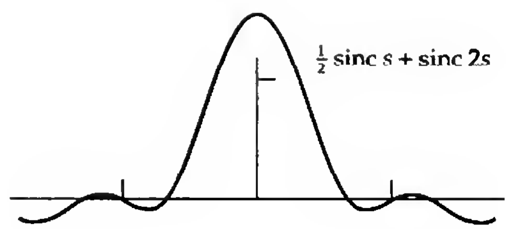

$\cos^2 \pi x \Pi(x)$

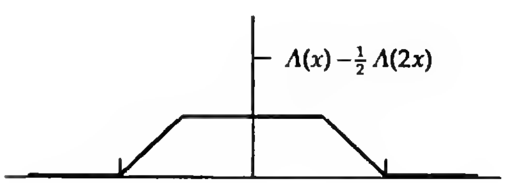

$\frac{1}{4}\operatorname{sinc}(s+1) + \frac{1}{2}\operatorname{sinc} s + \frac{1}{4}\operatorname{sinc}(s-1)$

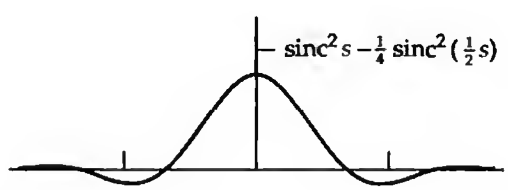

$\frac{1}{2}\cos \pi x \Pi(x) + \frac{1}{2}\Pi(x)$


$\frac{1}{4}\operatorname{sinc}(s+\frac{1}{2}) + \frac{1}{2}\operatorname{sinc} s + \frac{1}{4}\operatorname{sinc}(s-\frac{1}{2})$

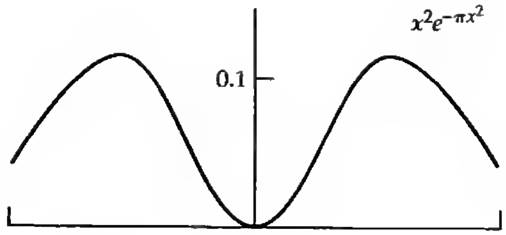

$\frac{1}{2}\Pi(\frac{1}{2}x) + \frac{1}{2}\Pi(x)$


$\frac{1}{2}\operatorname{sinc} s + \operatorname{sinc} 2s$

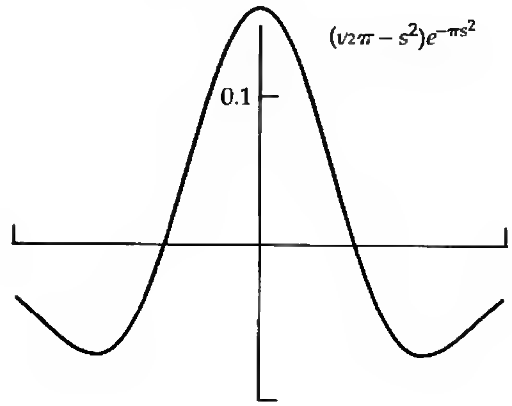

$\Lambda(x) - \frac{1}{2}\Lambda(2x)$


$\operatorname{sinc}^2 s - \frac{1}{4}\operatorname{sinc}^2(\frac{1}{2}s)$

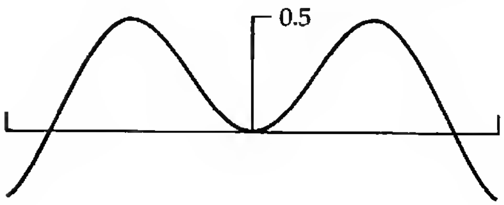
$x^2e^{-\pi x^2}$



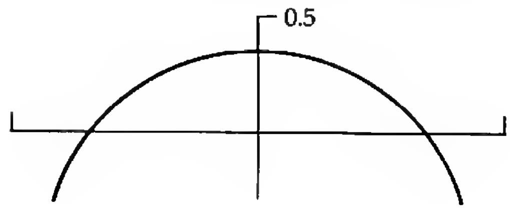
$(1/2\pi - s^2)e^{-\pi s^2}$



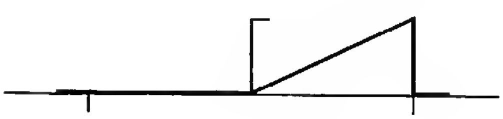
$J_2(2\pi x)$



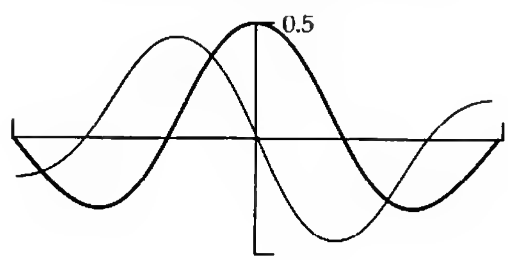
$(1-2s^2)/(1-s^2)^{\frac{1}{2}}\Pi(s/2)$



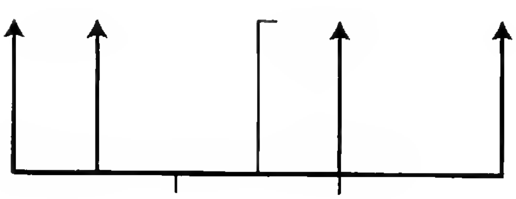
$x\Pi(x-\frac{1}{2})$



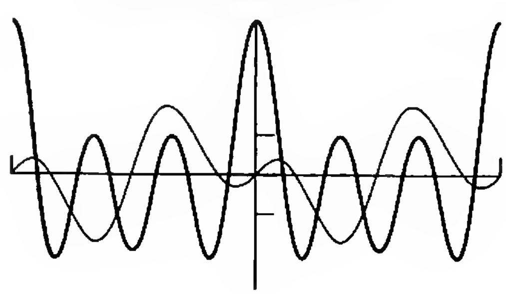
$\mathrm{Re} = (\pi s \sin 2\pi s - \sin^2 \pi s)/2\pi s^2$
$\mathrm{Im} = (\pi s \cos 2\pi s - \frac{1}{2}\sin 2\pi s)/2\pi s^2$

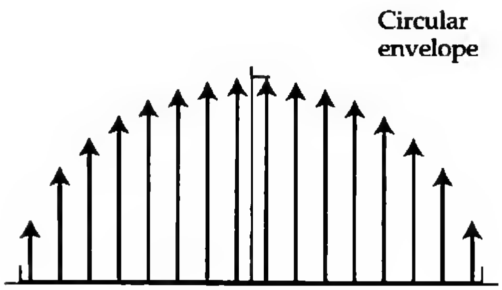


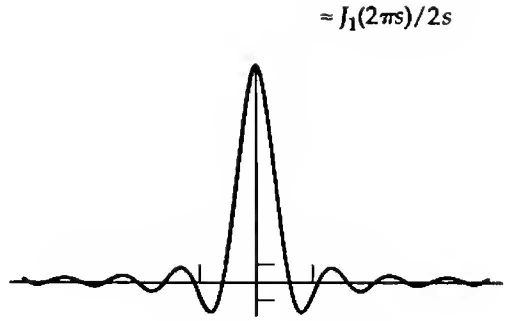
$\delta(x+3) + \delta(x+2) + \delta(x-1) + \delta(x-3)$

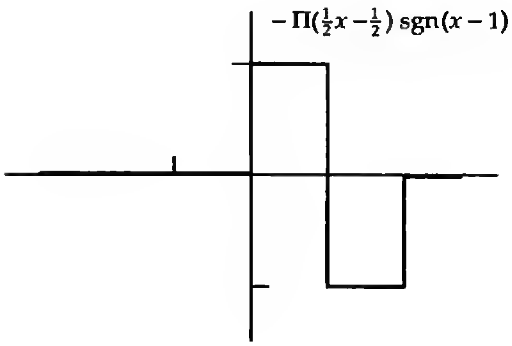


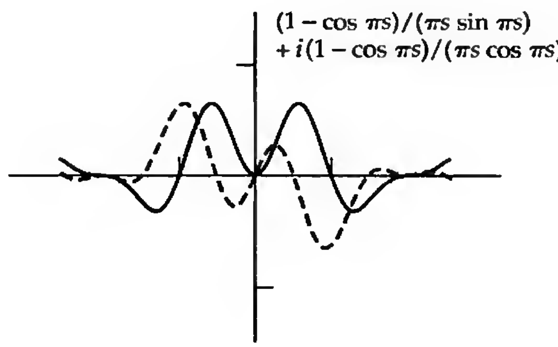
$\mathrm{Re} = 2\cos 6\pi s + \cos 4\pi s + \cos 2\pi s$
$\mathrm{Im} = \sin 4\pi s - \sin 2\pi s$



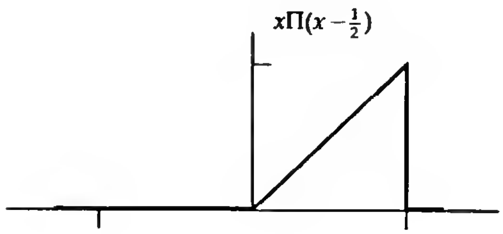

Circular
envelope

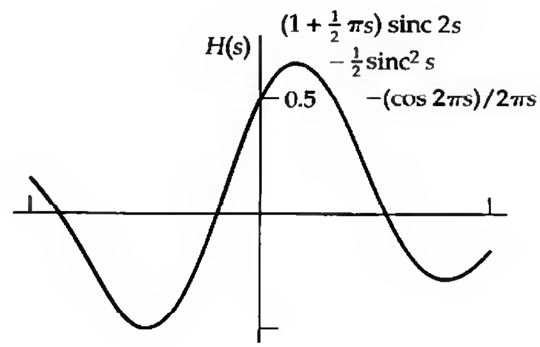

$\approx J_1(2\pi s)/2s$


$-\Pi(\frac{1}{2}x-\frac{1}{2})\,\mathrm{sgn}(x-1)$

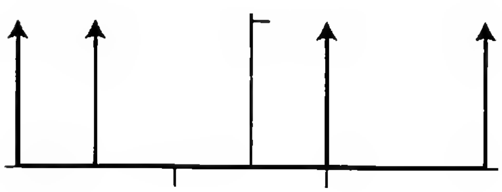

$(1-\cos\pi s)/(\pi s\sin\pi s)$
$+i(1-\cos\pi s)/(\pi s\cos\pi s)$

HARTLEY TRANSFORMS OF SOME FUNCTIONS WITHOUT SYMMETRY

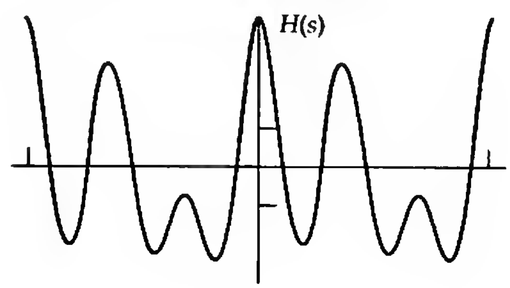




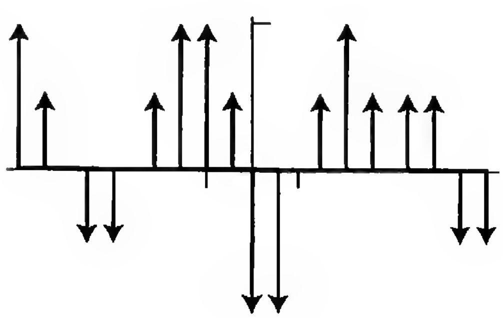
$\delta(x+3)+\delta(x+2)+\delta(x-1)+\delta(x-3)$

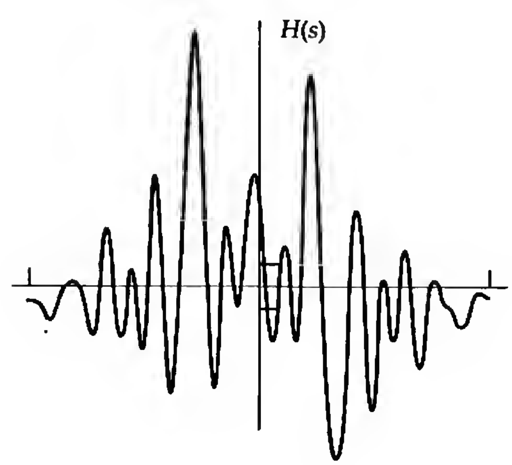


$2\cos 6\pi s + \operatorname{cas}(-4\pi s) + \operatorname{cas} 2\pi s$
where $\operatorname{cas} s = \cos s + \sin s$

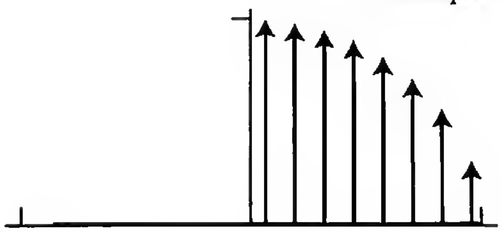


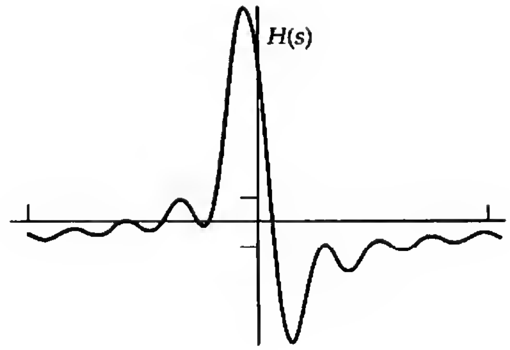
Correlated impulse sequence





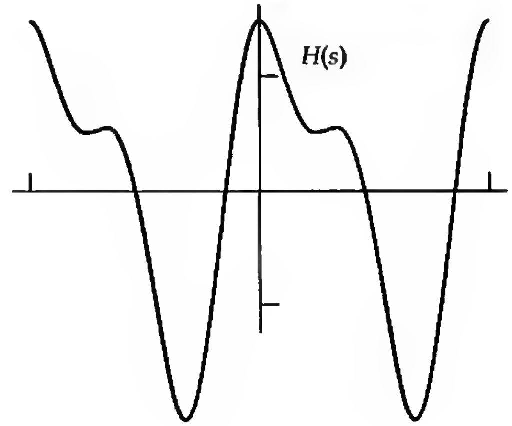
Circular envelope



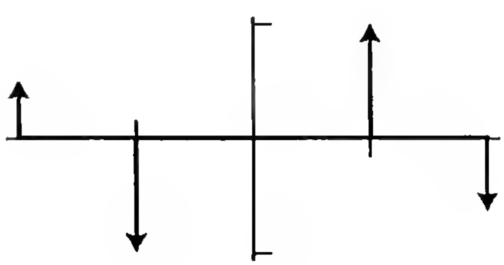


$\frac{1}{2}\delta(x+2)+\delta(x-1)$

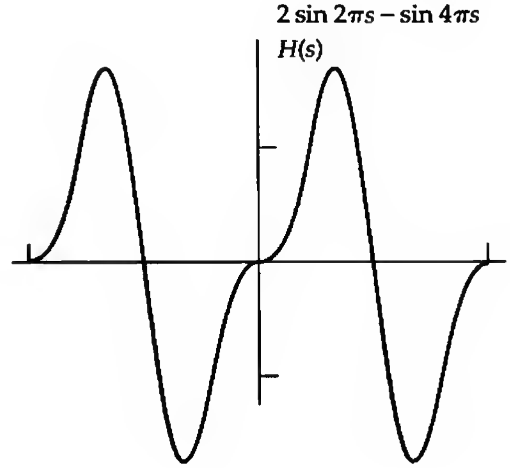
$\frac{1}{2}\operatorname{cas}(-4\pi s)+\operatorname{cas}(2\pi s)$



$\frac{1}{2}\delta(x+2)-\delta(x+1)+\delta(x-1)-\frac{1}{2}\delta(x-2)$



$2\sin 2\pi s-\sin 4\pi s$



23

The Life of Joseph Fourier

Baron Jean-Baptiste-Joseph Fourier (March 21, 1768–May 16, 1830), born in poor circumstances in Auxerre, introduced the idea that an arbitrary function, even one defined by different analytic expressions in adjacent segments of its range (such as a staircase waveform), could nevertheless be represented by a single analytic expression. This idea encountered resistance at the time but has proved to be central to many later developments in mathematics, science, and engineering. It is at the heart of the electrical engineering curriculum today. Fourier came upon his idea in connection with the problem of the flow of heat in solid bodies, including the earth.

The formula

$$\tfrac{1}{2}x = \sin x - \tfrac{1}{2}\sin 2x + \tfrac{1}{3}\sin 3x + \dots$$

was published by Leonhard Euler (1707–1783) before Fourier's work began, so you might like to ponder the question why Euler did not receive the credit for Fourier's series.

Fourier was obsessed with heat, keeping his rooms uncomfortably hot for visitors, while also wearing a heavy coat himself. Some traced this eccentricity back to his 3 years in Egypt, where he went in 1798 with the 165 savants on Napoleon's expedition to civilize the country.

Prior to the expedition Fourier was a simple professor of mathematics, but he now assumed administrative duties as secretary of the Institut d'Egypte, a scientific body that met in the harem of the palace of the Beys. Fourier worked on the theory of equations at this time but his competence at administration led to political and diplomatic assignments that he also discharged with success. It should be recalled that the ambitious studies in geography, archaeology, medicine, agriculture, natural history, and so on, were being carried out at a time when Napoleon was fighting Syrians in Palestine, repelling Turkish invasions, hunting Murad Bey, the elusive Mameluke chief, and all this without support of his fleet, which had

been obliterated by Nelson at the Battle of the Nile immediately after the disembarkation.

Shortly before the military capitulation in 1801, the French scientists put to sea but were promptly captured with all their records by Sidney Smith, commander of the British fleet. However, in accordance with the gentlemanly spirit of those days, Smith put the men ashore, retained the documents and collections for safekeeping, and ultimately delivered the material to Paris in person, except for the Rosetta stone, key to Egyptian hieroglyphics, which stands today in the British Museum memorializing both Napoleon's launching of Egyptology and his military failure.

The English physicist Thomas Young (1773–1829), father of linearity, is well known for establishing the transverse wave nature of light, explaining polarization, and also for introducing the double pinhole interferometer, which exhibits the Fourier analysis of an optical object. Less well known is that he shared an interest in Egyptology with Fourier: he worked on the Rosetta stone, explained the demotic and hieratic scripts as descended from hieroglyphic writing, and isolated and identified consonantal signs.

Fourier was appointed as Prefect of Isère by Napoleon in 1802 after a brief return to his former position as Professor of Analysis at the Ecole Polytechnique in Paris. His duties in Grenoble included taxation, military recruiting, enforcing laws, and carrying out instructions from Paris and writing reports. He soothed the wounds remaining from the Revolution of 1789, drained 80,000 km^2 of malarial swamps, and built the French section of the road to Torino.

By 1807, despite official duties, Fourier had written down his theory of heat conduction, which depended on the essential idea of analyzing the temperature distribution into spatially sinusoidal components; but doubts expressed by Laplace and Lagrange hindered publication. Criticisms were also made by Biot and Poisson. Even so, the Institut set the propagation of heat in solid bodies as the topic for the prize in mathematics for 1811, and the prize was granted to Fourier but with a citation mentioning lack of generality and rigor. The fact that publication was then further delayed until 1815 can be seen as an indication of the deep uneasiness about Fourier analysis that was felt by the great mathematicians of the day.

It is true that the one-dimensional distribution of heat in a straight bar would require a Fourier integral for its correct expression. Fourier avoided this complication by considering heat flow in a ring, that is, a bar that has been bent into a circle. In this way, the temperature distribution is forced to be spatially periodic. There is essentially no loss of generality because the circumference of the ring can be supposed larger than the greatest distance that could be of physical interest on a straight bar conducting heat. This idea of Fourier remains familiar as one of the textbook methods of approaching the Fourier integral as a limit, starting from a Fourier series representation.

Fourier was placed in a tricky position in 1814, when Napoleon abdicated and set out for Elba with every likelihood of passing southward through Grenoble, on what has come to be known today as the Route Napoléon. To greet his old mas-

ter would jeopardize his standing with the new king, Louis XVIII, who in any case might not look favorably on old associates and appointees of the departing emperor. Fourier influenced the choice of a changed route and kept his job. But the next year Napoleon reappeared in France, this time marching north through Grenoble where he fired Fourier, who had made himself scarce. Nevertheless, three days later Fourier was appointed Prefect of the Rhône at Lyons, thus surviving two changes of régime. Of course, only 100 days elapsed before the king was back in control and Napoleon was on his way to the south Atlantic, never to return. Fourier's days in provincial government then ended and he moved to Paris to enter a life of science and scientific administration, being elected to the Académie des Sciences in 1817, to the position of permanent secretary in 1823, and to the Académie Française in 1826. He never married.

At the beginning mention was made of Euler's formula. The formula is correct for $-\pi < x < \pi$ but not for other ranges of x. The RHS is the Fourier series for the sawtooth periodic function $(x/2)\Pi(x/2\pi) * (2\pi)^{-1}\text{III}(x/2\pi)$. Fourier wrote, around 1808–1809, "The equation is no longer true when the value of x is between π and 2π. However, the second side of the equation is still a convergent series but the sum is not equal to $x/2$. Euler, who knew this equation, gave it without comment." (Quotation from J. Hérivel, *Joseph Fourier, the Man and the Physicist*, Clarendon Press, Oxford, 1975, p. 319.)

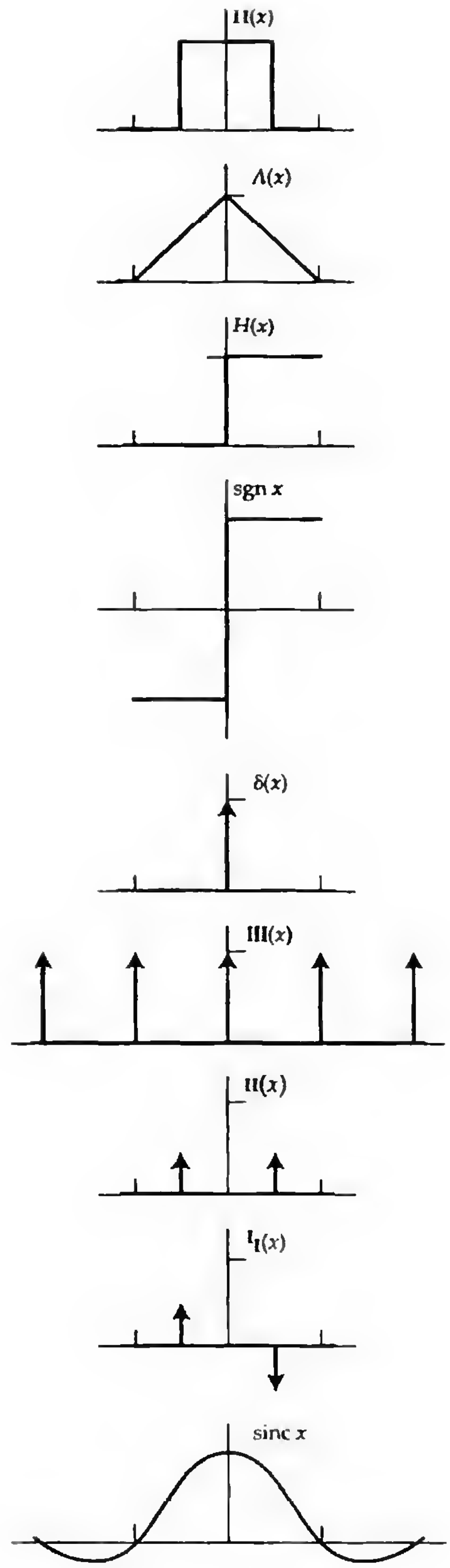
Index

Pages in *italics* refer to end-of-chapter bibliographies.

Special symbols

Function	Notation
Rectangle	$\Pi(x) = \begin{cases} 1 & \lvert x\rvert < \frac{1}{2} \\ 0 & \lvert x\rvert > \frac{1}{2} \end{cases}$
Triangle	$\Lambda(x) = \begin{cases} 1 - \lvert x\rvert & \lvert x\rvert < 1 \\ 0 & \lvert x\rvert > 1 \end{cases}$
Heaviside unit step	$H(x) = \begin{cases} 1 & x > 0 \\ 0 & x < 0 \end{cases}$
Sign (signum)	$\operatorname{sgn} x = \begin{cases} 1 & x > 0 \\ -1 & x < 0 \end{cases}$
Impulse symbol	$\delta(x)$
Sampling or replicating symbol	$\text{III}(x) = \sum_{-\infty}^{\infty} \delta(x - n)$
Even impulse pair	$\text{II}(x) = \frac{1}{2}\delta(x + \frac{1}{2}) + \frac{1}{2}\delta(x - \frac{1}{2})$
Odd impulse pair	${}^{\text{I}}_{\text{I}}(x) = \frac{1}{2}\delta(x + \frac{1}{2}) - \frac{1}{2}\delta(x - \frac{1}{2})$
Filtering or interpolating	$\operatorname{sinc} x = \frac{\sin \pi x}{\pi x}$, $\operatorname{jinc} r = \frac{J_1(\pi r)}{2r}$
Asterisk notation for convolution	$f(x) * g(x) \triangleq \int_{-\infty}^{\infty} f(u) g(x - u)\, du$
Asterisk notation for serial products	$\{f_i\} * \{g_i\} \triangleq \left\{ \sum_j f_j g_{i-j} \right\}$
Pentagram notation	$f(x) \star g(x) \triangleq \int_{-\infty}^{\infty} f(u) g(x + u)\, du$
Various two-dimensional functions	${}^2\Pi(x,y) = \Pi(x)\Pi(y)$ ${}^2\delta(x,y) = \delta(x)\,\delta(y)$ ${}^2\text{III}(x,y) = \text{III}(x)\text{III}(y)$ ${}^2\operatorname{sinc}(x,y) = \operatorname{sinc} x \operatorname{sinc} y$

Guide to reference data


Π(x)
Λ(x)
H(x)
sgn x
δ(x)
III(x)
II(x)
I_I(x)
sinc x